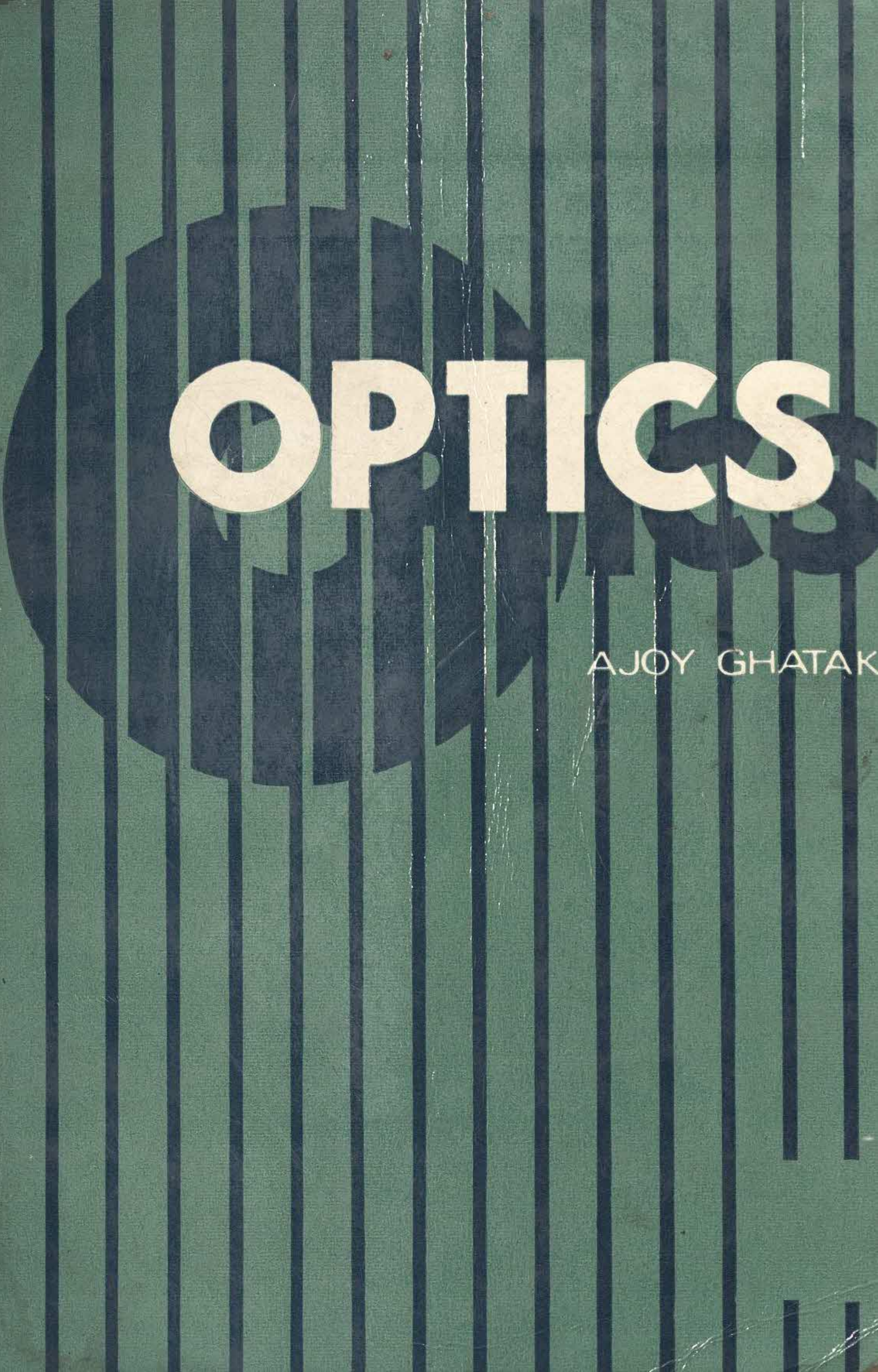


OPTICS

AJOY GHATAK

OPTICS

The completest knowledge of the laws of nature does not carry with it the power of prediction, nor of mastery over Nature. If the universe is a machine its levers and wheels are too fine for our hands to manipulate. We can learn and guide its large scale motions only. Beneath our veiled sight it quivers its eternal quest.

Max Born

McGraw-Hill Offices :

New Delhi
New York
St Louis
San Francisco
Auckland
Bogotá
Guatemala
Hamburg
Lisbon
London
Madrid
Mexico
Milan
Montreal
Panama
Paris
San Juan
São Paulo
Singapore
Sydney
Tokyo
Toronto

OPTICS

AJOY GHATAK

Department of Physics
Indian Institute of Technology
New Delhi

S.C.E.R.T., West Bengal
Library
Calcutta
25/3, B. C. Road

TATA McGRAW-HILL PUBLISHING COMPANY LIMITED
NEW DELHI

535
GHA

S.C.E.R.T., West Bengal
Date 15.3.91
Acc. No 4958

© 1977, Tata McGraw-Hill Publishing Company Limited

Thirteenth Reprint 1989

No part of this publication may be reproduced in any form or by any means
without the prior written permission of the publishers

This edition can be exported from India only by the publishers,
Tata McGraw-Hill Publishing Company Limited

Published by Tata McGraw-Hill Publishing Company Limited,
4/12 Asaf Ali Road, New Delhi-110002, and printed at
Pushp Print Services, Moujpur, Delhi 110 053

to my students

Preface

In recent years, the field of optics has become extremely important because of its applications in many diverse areas. This is primarily due to the advent of the laser which has helped in the practical realization of many interesting experiments and has also opened up new fields of activities. Although there exist a large number of texts on optics, most of them do not discuss some of its modern aspects like coherence, holography, self-focusing, spatial frequency filtering, physics of lasers, etc. The present book gives a balanced account of traditional optics as well as some of the recent developments in this field.

The book has grown out of the lectures delivered by the author to the students at the Indian Institute of Technology, New Delhi. Care has been taken to develop the topics from the first principles so that the book can also be used for self study; the underlying physical principles have also been emphasized.

Some special features of the book are given below:

(1) Chapter 1, in addition to a brief historical introduction to the field of optics, tries to give a detailed discussion on the nature of light.

(2) Chapter 2 studies the propagation of rays in inhomogeneous media and gives a fairly complete account of the phenomena of mirage and looming. Fermat's principle has also been used to study the refraction of rays at the interface of an isotropic and an anisotropic medium.

(3) It is now well known that, in addition to its elegance, the matrix method is a very convenient approach for the tracing of rays through complex optical systems; the third chapter develops this method to be used to study

the image formation in simple optical systems. The matrix method has also been used in chapter 22 to study the resonator systems employed in lasers.

(4) In chapter 6 expressions for the refractive index have been obtained from the first principles. The dependence of the refractive index on the wavelength has also been discussed. The fact that the ionosphere reflects electromagnetic waves of low frequencies and transmits electromagnetic waves of high frequencies has also been explained.

(5) In chapter 7, the Fourier integral theorem has been used to study the propagation and spreading of a wave packet. The actual spreading of a Gaussian wave packet in a dispersive medium has also been discussed.

(6) Chapter 12 presents a detailed discussion on the localization of interference fringes.

(7) The concepts of spatial and temporal coherence have been developed very carefully in chapter 14 and the results of the recent experiments on optical beats have also been given.

(8) Chapter 15, in addition to a detailed discussion on Fraunhofer diffraction, also discusses X-ray diffraction, self-focusing phenomenon and spatial frequency filtering; the latter two fields have received considerable attention after the development of the laser.

(9) In chapter 16, in addition to a fairly rigorous treatment of Fresnel diffraction, a careful analysis of the transition from the Fresnel region to the Fraunhofer region is also made.

(10) In chapter 17, the basic principle of holography and some of its applications have been discussed.

(11) In chapter 18, the polarization property of the electromagnetic waves has been discussed.

(12) In chapters 19 and 20, the solutions of Maxwell's equations are studied from which the laws of reflection and refraction of plane electromagnetic waves by a plane dielectric and a plane metallic surface are derived. It has been pointed out that it is the frequency of the electromagnetic wave which determines whether a particular material would behave as a conductor or a dielectric. Chapter 20 also discusses evanescent waves which are generated when a beam undergoes total internal reflection.

(13) In chapter 21, the particle aspect of electromagnetic radiation has been discussed.

(14) In the last chapter, the basic physics of a laser has been discussed in sufficient detail.

A large number of problems have been appended to each chapter and also references to articles and books for additional reading.

Although some of the topics discussed are somewhat of an advanced nature, they have been dealt with in a way by which even an undergraduate student should be able to comprehend them.

In a book of over 500 pages, errors are bound to creep in. The author will be grateful if these are pointed out to him.

AJOY GHATAK

Acknowledgements

I have dedicated this book to my students who have incessantly helped me during its preparation. Although all my students have helped me in one way or the other, I owe a special debt to K. Thyagarajan for going through the entire text, making numerous suggestions and also helping me in writing some of the chapters. Special thanks are also due to Enakshi Khular, Aruna Rohra and Anurag Sharma who have made valuable suggestions, and have helped me in the preparation of the book and in carrying out many of the numerical calculations. I am also grateful to Anjana Gupta, B.D. Gupta, Arun Kumar, Sugata Mitra and R.P. Sharma for their help in the preparation of the book.

Portions of this book have been used for teaching the undergraduate students of the Indian Institute of Technology, New Delhi. My thanks are due to Dr. I.C. Goyal, Dr. S.C. Mathur, Dr. S.S. Mathur, Dr. R.K. Puri, Dr. D.P. Tewari and Dr. B.B. Tripathi for their suggestions and constructive criticisms. I am also grateful to Dr. R.S. Sirohi of the I.I.T. Madras and Dr. Vishwamitter of the Punjab Agricultural University, Ludhiana, for valuable suggestions.

I am indebted to Academic Press, Inc., Addison Wesley Publishing Co., American Association of Physics Teachers, Educational Development Center (USA), McGraw-Hill Book Co. and Oxford University Press, for their kind permissions to use material from their publications. I am grateful to Dr. R.S. Sirohi, Dr. R.A. Phillips, Dr. K. Thyagarajan, Mr. K.K. Gupta, Mr. Anurag Sharma, Mr. G. Bose and Educational Development Center

(USA) for providing me with the glossy prints of some of the photographs appearing in the book.

Thanks are also due to Mr. T.N. Gupta, Mr. N.S. Gupta and Mr. K. Giri for their help in the preparation of the manuscript.

I am particularly grateful to Professor M.S. Sodha for his constant encouragement and interest in this endeavour. Finally, my special thanks to my wife Gopa, for her patience and understanding.

Contents

Preface vii

Acknowledgements ix

1. What is Light? 1

1.1 Introduction 1
1.2 The Corpuscular Model 2
1.3 The Wave Model 3
1.4 The Particle Nature of Radiation 7
1.5 The Uncertainty Principle 8
1.6 The Probabilistic Interpretation of Matter Waves 10
1.7 An Understanding of Interference Experiments 12
1.8 The Polarization of a Photon 14
Problems 16
References and Suggested Reading 17

PART I GEOMETRICAL OPTICS

2. Fermat's Principle and its Applications 21

2.1 Introduction 21
2.2 Laws of Reflection and Refraction from Fermat's Principle 23
2.3 Application of Fermat's Principle to Inhomogeneous Media 29
2.4 Refraction of Rays at the Interface between an Isotropic and an Anisotropic Medium

Problems 41
References and Suggested Reading 44

3. Matrix Method in Paraxial Optics 45

- 3.1 Introduction 45
- 3.2 The Matrix Method 47
- 3.3 Unit Planes 55
- 3.4 Nodal Planes 58
- 3.5 A System of Two Thin Lenses 59

Problems 63
References and Suggested Reading 64

4. Aberrations 65

- 4.1 Introduction 65
- 4.2 Chromatic Aberration 66
  - 4.2.1 The Achromatic Doublet 67
  - 4.2.2 Removal of Chromatic Aberration of a Separated Doublet 69
- 4.3 Spherical Aberration 70
- 4.4 Coma 81
- 4.5 Astigmatism and Curvature of Field 84
- 4.6 Distortion 87

Problems 88
References and Suggested Reading 88

PART II VIBRATION AND WAVES

5. Simple Harmonic Motion 93

- 5.1 Definition 93
- 5.2 Examples of Simple Harmonic Motion 97

Problems 101
References and Suggested Reading 102

6. Forced Vibrations and Origin of Refractive Index 103

- 6.1 Damped Simple Harmonic Motion 103
- 6.2 Forced Vibrations 107
- 6.3 Resonance 108
- 6.4 Origin of Refractive Index 111

Problems 118
References and Suggested Reading 120

7. Fourier Series and Applications 121

- 7.1 Fourier Analysis 121
- 7.2 Transverse Vibrations of a Plucked String 126
- 7.3 Application of Fourier Series in Forced Vibrations 130

7.4 The Fourier Integral 132
7.5 Group Velocity of a Wave Packet 134
Problems 142
References and Suggested Reading 143

8. Wave Propagation and the Wave Equation 144

8.1 Introduction 144
8.2 Sinusoidal Waves : Concept of Frequency and Wavelength 147
8.3 Types of Waves 150
8.4 Energy Transport in Wave Motion 150
8.5 The One-Dimensional Wave Equation 152
8.6 Transverse Vibrations of a Stretched String 154
8.7 Longitudinal Sound Waves in a Solid 155
8.8 Longitudinal Waves in a Gas 158
8.9 The General Solution of the One-Dimensional Wave Equation 160
Problems 168
References and Suggested Reading 169

9. Huygens' Principle and its Applications 170

9.1 Introduction 170
9.2 Huygens' Theory 171
9.3 Rectilinear Propagation 173
9.4 Application of Huygens' Principle to Study Refraction and Reflection 174
9.5 Huygens' Principle in Inhomogeneous Media 185
Problems 186
References and Suggested Reading 187

PART III INTERFERENCE

10. Superposition of Waves 191

10.1 Introduction 191
10.2 Stationary Waves on a String 192
10.3 Stationary Waves in a String whose Ends are Fixed 195
10.4 Stationary Light Waves: Ives' and Wiener's Experiments 197
10.5 Superposition of Two Sinusoidal Waves 198
10.6 The Graphical Method for Studying Superposition of Sinusoidal Waves 200
10.7 The Complex Representation 203
Problems 204
References and Suggested Reading 205

11. Two Beam Interference by Division of Wavefront 206

11.1 Introduction 206

11.2 Interference Pattern Produced on the Surface of Water 207
11.3 Coherence 212
11.4 Interference of Light Waves 214
11.5 The Interference Pattern 216
11.6 The Intensity Distribution 218
11.7 Fresnel's Two Mirror Arrangement 223
11.8 Fresnel Biprism 224
11.9 Interference with White Light 226
11.10 Displacement of Fringes 227
11.11 The Lloyd's Mirror Arrangement 228
11.12 Phase Change on Reflection 229
Problems 231
References and Suggested Reading 233

12. Interference by Division of Amplitude 234

12.1 Introduction 234
12.2 Interference by a Plane Parallel Film when Illuminated by a Plane Wave 234
12.3 The Cosine Law 238
12.4 Non Reflecting Films 239
12.5 High Reflectivity by Thin Film Deposition 242
12.6 Interference by a Plane Parallel Film when Illuminated by a Point Source 243
12.7 Interference by a Film with Two Non-Parallel Reflecting Surfaces 247
12.8 Colours of Thin Films 253
12.9 Newton's Rings 253
12.10 The Michelson Interferometer 260
Problems 264
References and Suggested Reading 265

13. Multiple Beam Interferometry 266

13.1 Introduction 266
13.2 Multiple Reflections from a Plane Parallel Film 266
13.3 Fabry-Perot Interferometer 271
    13.3.1 Resolving Power of a Fabry-Perot Interferometer 272
13.4 The Lummer-Gehrcke Plate 275
13.5 Interference Filters 277
Problems 278
References and Suggested Reading 279

14. Coherence 280

14.1 Temporal Coherence 280
14.2 The Line Width 282
14.3 Fringe Visibility in the Double Hole Experiment 286
14.4 Spatial Coherence 289

14.5 Michelson's Stellar Interferometer 292
14.6 Optical Beats 293
Problems 297
References and Suggested Reading 298

PART IV DIFFRACTION

15. Fraunhofer Diffraction 301

15.1 Introduction 301
15.2 Single Slit Diffraction Pattern 303
 15.2.1 The Positions of Maxima and Minima 305
15.3 Diffraction by a Circular Aperture 311
15.4 Limit of Resolution 313
15.5 Two Slit Fraunhofer Diffraction Pattern 317
 15.5.1 Positions of Maxima and Minima 319
15.6 N-Slit Fraunhofer Diffraction Pattern 321
 15.6.1 Positions of Maxima and Minima 323
 15.6.2 Width of the Principal Maximum 325
15.7 The Diffraction Grating 325
 15.7.1 The Grating Spectrum 326
 15.7.2 Resolving Power of a Grating 329
15.8 Oblique Incidence 330
15.9 X-ray Diffraction 331
15.10 The Self-Focussing Phenomenon 338
15.11 Spatial Frequency Filtering 343
Problems 347
References and Suggested Reading 348

16. Fresnel Diffraction 349

16.1 Introduction 349
16.2 Fresnel Half-Period Zones 350
16.3 The Zone-plate 355
16.4 Diffraction at a Straight Edge 358
16.5 Diffraction of a Plane Wave by a Straight Edge : A More Rigorous Approach 360
16.6 Diffraction of a Plane Wave by a Long Narrow Slit and Transition to the Fraunhofer Region 368
16.7 Fresnel Diffraction by a Circular Aperture 373
16.8 Diffraction by a Circular Disc 376
16.9 Fraunhofer Diffraction by an Aperture 377
 16.9.1 Rectangular Aperture 378
 16.9.2 Circular Aperture 379
Problems 382
References and Suggested Reading 383

17. Holography 385

17.1 Introduction 385
17.2 Theory 387
17.3 Requirements 394
17.4 Some Applications 394
Problems 397
References and Suggested Reading 397

PART V ELECTROMAGNETIC CHARACTER OF LIGHT

18. Polarization and Double Refraction 401

18.1 Introduction 401
18.2 Production of Polarized Light 406
18.2.1 The Wire Grid Polarizer and the Polaroid 407
18.2.2 Polarization by Reflection 408
18.2.3 Polarization by Double Refraction 409
18.2.4 Polarization by Scattering 410
18.3 Malus' Law 412
18.4 Superposition of Two Disturbances 412
18.4.1 The Mathematical Analysis 417
18.5 The Phenomenon of Double Refraction 418
18.6 Interference of Polarized Light 424
18.7 Analysis of Polarized Light 429
18.8 Plane Wave Propagation in Anisotropic Media 430
Problems 437
References and Suggested Reading 437

19. Electromagnetic Waves 440

19.1 Maxwell's Equations 440
19.2 Plane Waves in a Dielectric 441
19.3 The Three-Dimensional Wave Equation in a Dielectric 447
19.4 The Poynting Vector 449
19.4.1 The Oscillating Dipole 451
19.5 Energy Density and Intensity of an Electromagnetic Wave 453
19.6 Radiation Pressure 454
19.7 The Wave Equation in a Conducting Medium 457
19.8 The Continuity Conditions 459
19.9 Physical Significance of Maxwell's Equations 461
Problems 464
References and Suggested Reading 466

20. Reflection and Refraction of Electromagnetic Waves 467

20.1 Introduction 467
20.2 Reflection at an Interface of Two Dielectrics 467

20.3 Reflection by a Conducting Medium 486
Problems 488
References and Suggested Reading 489

PART VI PHOTONS

21. The Particle Nature of Radiation 493

21.1 Introduction 493
21.2 The Photoelectric Effect 493
21.3 The Compton Effect 497
Problems 502
References and Suggested Reading 502

PART VII SOURCES OF COHERENT LIGHT

22. Lasers : An Elementary Account 507

22.1 Introduction 507
22.2 Spontaneous and Stimulated Emission 510
22.3 Population Inversion 512
22.4 Resonators 517
22.5 Helium-Neon Laser 524
Problems 525
References and Suggested Reading 525

General Books on Optics 528

Author Index 531
Subject Index 534

List of Plates

Plate *Facing Page*

1 Photographs showing the quality of a picture obtainable from various numbers of photons. (a), (b), (c), (d), (e) and (f) correspond to 3×10^3 photons, 1.2×10^4 photons, 9.3×10^4 photons, 7.6×10^5 photons, 3.6×10^6 photons and 2.8×10^7 photons respectively. (From A. Rose, "Quantum Effects in Human Vision", *Advances in Biological and Medical Physics*, Vol. V, Academic Press, 1957; used with permission.) 11

2 The spherical aberration of a convex lens. (Photograph courtesy Mr. K.K. Gupta.) 72

3 Image of a point source showing coma. (After H.F. Meiners, *Physics Demonstration Experiments*, Vol. II, The Ronald Press Co., New York, 1970; used with permission.) 83

4 Diffraction of straight water waves when they pass through an opening. (After PSSC PHYSICS, D.C. Heath & Co., Boston, Mass., 1965; used with permission.) 173

5 The actual interference pattern produced from two point sources vibrating in phase in a ripple tank. (After PSSC PHYSICS, D.C. Heath & Co., Boston, Mass., 1965; used with permission.) 209

6 Interference fringes produced by two slits using monochromatic light. (Photograph courtesy Dr. R.S. Sirohi.) 221

7 A typical fringe pattern produced by an air-film formed between two

Plate *Facing Page*

glass surfaces (which are not optically flat) and placed in contact with each other. Whenever the thickness of the air film is $m\lambda/2$, we obtain a dark fringe and when the thickness is $(m+\frac{1}{2})\lambda/2$, we obtain a bright fringe. Each fringe describes a locus of equal thickness of the film. (Photograph courtesy Dr. R.S. Sirohi.) 253

8 Newton's rings as observed in reflection. The rings observed with transmitted light are of much poorer contrast. (Photograph courtesy Mr. G. Bose.) 255

9 Fringes observed using (a) Michelson interferometer (b) Febry-Perot interferometer. (From F.A. Jenkins and H.E. White, *Fundamentals of Optics*, McGraw-Hill Book Co., New York, 1976; used with permission.) 272

10 The single slit diffraction patterns corresponding to $b=0.0088$, 0.0176, 0.035 and 0.070 cm respectively. The wavelength of the light used is 6.328×10^{-5} cm. (After H.F. Meiners, *Physics Demonstration Experiments*, Volume II, The Ronald Press Co., New York, 1970; used with permission.) 309

11 The Airy pattern. (Photograph courtesy Mr. Anurag Sharma.) 311

12 The actual diffraction patterns corresponding to two point sources when they are (a) well resolved (b) just resolved. (Photographs courtesy Mr. Anurag Sharma.) 315

13 The double slit Fraunhofer diffraction pattern corresponding to $b=0.0088$ cm and $\lambda=6.328\times10^{-5}$ cm. The values of d are 0, 0.0176, 0.035 and 0.070 cm respectively. (After H.F. Meiners, *Physics Demonstration Experiments*, Volume II, The Ronald Press Co., New York, 1970; used with permission.) 319

14 The multiple-slit Fraunhofer diffraction patterns corresponding to $b=0.0044$ cm, $d=0.0132$ cm and $\lambda=6.328\times10^{-5}$ cm. The number of slits are 1, 2, 3 and 4 respectively. (After H.F. Meiners, *Physics Demonstration Experiments*, Volume II, The Ronald Press Co., New York, 1970; used with permission.) 324

15 (a) Shows a photograph consisting of regularly spaced black and white squares of varying sizes. When a pinhole is placed in the Fourier transform plane to block the high frequency components an image of the form shown in (c) is obtained; the frequency spectrum is shown in (b). Notice that in (c) shades of grey appear as well as details such as the missing part of the eye glass frame. (After R.A. Phillips, "Spatial Filtering Experiments for Undergraduate Laboratories", *American Journal of Physics*, 37, 536, 1969.) 346

16 The straight edge diffraction pattern. (Photograph courtesy Dr. K. Thyagarajan.) 366

17 Fraunhofer diffraction pattern of a square aperture. (Photograph courtesy Mr. K.K. Gupta.) 379

Plate *Facing Page*

18 (a) An ordinary photograph of an object. (b) The hologram of the object produced by a method similar to one as shown in Fig. 17.1. (c) The reconstructed image as seen by an observer. (d) A magnified view of a small portion of the hologram shown in (b). (Photographs courtesy Dr. R.S. Sirohi.) 386

19 Interference fringes produced in the measurement of Young's modulus using double exposure interferometry. (Photograph courtesy Dr. R.S. Sirohi.) 397

1 What is Light?

How little do we know, that which we are!
How less, what we may be!

BYRON

1.1 Introduction

We begin this chapter by quoting Feynman:*

"Newton thought that light was made up of particles, but then it was discovered that it behaves like a wave. Later, however (in the beginning of the twentieth century), it was found that light did indeed sometimes behave like a particle. Historically, the electron, for example, was thought to behave like a particle, and then it was found that in many respects it behaved like a wave. So it really behaves like neither. Now we have given up. We say: 'It is like *neither*.'

There is one lucky break, however, electrons behave just like light. The quantum behaviour of atomic objects (electrons, protons, neutrons, photons, and so on) is the same for all, they are all 'particle-waves', or whatever you want to call them."

The important thing to note here is the fact that light is neither a wave nor a particle. The modern quantum theory describes it in a sort of abstract way which cannot be connected with everyday experience.

In this chapter, we will make a brief historical survey of the important experiments which led to models regarding the nature of light and near the end of the chapter, we will qualitatively discuss how the wave and the particle aspects of radiation can be explained on the basis of the uncertainty principle and the probabilistic interpretation of matter waves.

**The Feynman Lectures on Physics*, Vol. III, by R.P. Feynman, R.B. Leighton and M. Sands, Addison Wesley Publishing Co., Reading, Massachusetts, 1965.

1.2 The Corpuscular Model

Perhaps the most important experimental fact which led to the early belief in the corpuscular model was the propagation of light through vacuum. It was thought that any wave phenomenon would require a material medium which would be responsible for the propagation of the wave. It was also believed that light travelled in straight lines which could be explained satisfactorily on the basis of the corpuscular model. Further, the laws of reflection and refraction* could also be explained on the basis of the corpuscular model of light. It may be mentioned that, contrary to popular belief, the corpuscular model of light is primarily due to Descartes rather than to Newton.** Here we would give the derivation of the Snell's law based on the corpuscular model.

In Fig. 1.1 we show the refraction of a particle at a plane surface ($y=0$); we are assuming that the motion is confined to the x-y plane. The trajectory of the particle is determined by the application of the laws of conservation of the energy E and the x component of the momentum p_x ($=p \sin \theta$) where θ is the angle that the direction of propagation makes with the y-axis. The conservation conditions lead to the following equations:

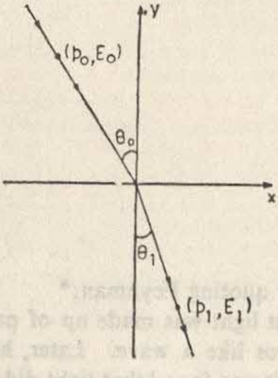


Fig. 1.1 Refraction of a corpuscle.

$$p_0 \sin \theta_0 = p_1 \sin \theta_1 \tag{1.1-1}$$

$$E_0 = E_1 \tag{1.1-2}$$

---

*The refraction law will be discussed in this secton; the reflection law immediately follows by considering the elastic reflection of a particle by a plane surface.

**For a good discussion of the corpuscular model see Joyce and Joyce, (1976); the discussion in this section follows this paper. These authors have also given an Eglish translation of the original derivation of Snell's law by Descartes; this derivation is equivalent to the corpuscular derivation which is attributed to Newton. [It may be of interest to mention that Newton (1642–1727) was only about eight years old when Descartes (1596–1650) died.] The law of refraction was discovered experimentally by Snell (1591–1626) in 1621, Descartes' derivation was published in 1637 and the first edition of Newton's *Optiks* was published in 1704.

Eq. (1.1-1) gives the Snell's law:

$$\frac{\sin\theta_0}{\sin\theta_1}=\frac{p_1}{p_0} \tag{1.1-3}$$

The angles θ_0 and θ_1 have been shown explicitly in Fig. 1.1. As discussed by Joyce and Joyce (1976), there exists a variety of relations between $\mathbf{p}$ and $\mathbf{v}$ and as such, Eq. (1.1-3) can assume a variety of forms. For example, if we consider only those particle-medium systems in which the mass ($=p/v$) has a constant value, then Eq. (1.1-3) becomes

$$\frac{\sin\theta_0}{\sin\theta_1}=\frac{v_1}{v_0} \quad \text{with} \quad E_0=E_1 \tag{1.1-4}$$

On the other hand, if we consider only non-dispersive systems, i.e. systems in which the group or particle speed* $v\left(=\frac{\partial E}{\partial p}=\frac{\partial\omega}{\partial k}\right)$ is equal to the phase speed c $\left(=\frac{E}{p}=\frac{\omega}{k}\right)$, then Eq.(1.1-3) reduces to

$$\frac{\sin\theta_0}{\sin\theta_1}=\frac{v_0}{v_1} \quad \text{with} \quad E_0=E_1 \tag{1.1-5}$$

One can immediately see that the value of $\frac{\sin\theta_0}{\sin\theta_1}$ as predicted by Eq. (1.1-4), is the reciprocal of the one predicted by Eq. (1.1-5). Equations (1.1-4) and (1.1-5) are special cases of Eq. (1.1-3) which is the physically correct form of the refraction law. Indeed, one can obtain different forms of the refraction law depending on the relation that one chooses between $\mathbf{p}$ and $\mathbf{v}$. Descartes and Newton conserved v_x rather than p_x, and obtained Eq. (1.1-4) which is physically correct in such special cases as a non-relativistic electrostatic electron lens where the final speed happens to be equal to v_0p_1/p_0.

On the other hand, Eq. (1.1-5) is the same as predicted by the Huygens' theory (see Sec. 9.4) and agrees with experimental observations for media which can be assumed to be non-dispersive. However, the equation is not, in general, valid for light because of dispersion. For further reading the reader is referred to the paper by Joyce and Joyce (1976), where the references to earlier studies on this topic are also cited.

1.3 The Wave Model

Although a corpuscular theory explains the propagation of light through free space and can be made to predict the correct forms of the laws of reflection and refraction, there exist a large number of experimental observations (like interference, diffraction, polarization, etc.), which cannot be explained on the assumption that light can be described as a classical particle. Historically, the "Newton's rings" which is a beautiful manifestation of the wave character of light,* was observed by Boyle (1627–1691) and by Hooke (1635–1703); the rings are named after Newton because he had given an explanation** of the formation of these rings by means of the corpuscular model, which was later found to be quite unsatisfactory.***

---

*The phase and the group velocities are discussed in Sec. 7.5. We should mention here that for light $E=h\nu=\hbar\omega$, where $\hbar=h/2\pi$, h is the Planck's constant and ν is the frequency; further $p=\frac{h}{\lambda}=\hbar k$ $\left(k=\frac{2\pi}{\lambda} \text{ and } \omega=2\pi\nu\right)$—see Sec. 1.4 and 1.5.

Grimaldi (1618–1663) was probably the first one to observe the diffraction of light, i.e. the presence of light in the geometrical shadow. Later, Hooke also observed this phenomenon. As will be discussed in later chapters, a satisfactory explanation of Newton's rings and of the diffraction phenomenon can only be given if one assumes a wave model of light. The wave model was first propounded by Huygens (1629–1695) around the later part of the seventeenth century. Using the wave model, Huygens could explain the laws of reflection and refraction (see Chapter 9) and he could also interpret the phenomenon of double refraction (see Chapter 18) discovered by Bartholinus (1625–1698). However, so compelling was Newton's authority that people around Newton had more faith on his corpuscular theory than Newton himself and no one really believed**** in Huygens' wave theory until the beginning of the nineteenth century when Young (1773–1829) performed the famous interference experiment (see Chapter 11) which could only be explained on assuming a wave model for light.***** In 1802, he gave a satisfactory explanation of the formation of Newton's rings. In 1808, Malus (1775–1812) observed the polarization of light but he did not try to interpret this phenomenon. Around 1816, Fresnel (1788–1827) gave a satisfactory explanation of the diffraction phenomenon by means of a wave theory and calculated the diffraction patterns produced by various types of apertures and edges. In 1816, Fresnel along with Arago (1786–1853) performed the famous experiment on the superposition of linearly polarized light waves which was explained by Young by assuming that light waves were transverse in character.

---

*We are not giving here the details of the experimental arrangement associated with Newton's rings; this is discussed in Sec. 12.9. Similarly, the phenomenon of diffraction, polarization and other characteristics associated with waves are discussed in detail in other chapters of this book.

**Newton's explanation of the rings has been discussed by many authors; see, e.g. Barton (1939).

***Further, when observed with white light, the rings are found to be coloured and Newton had revealed the basic quality of colour through his famous experiment of splitting the white light into its constituent colours by means of a prism.

****It should be pointed out that at the time of Huygens, light was thought to travel in straight lines and Huygens tried to invoke unrealistic assumptions in order to explain the rectilinear propagation of light using his wave theory. This drawback was one of the reasons for the immediate non-acceptance of the wave model.

*****Indeed, Young showed that the wavelength of the light waves was around 6×10^{-5} cm. Because of the smallness of the wavelength, the diffraction effects are small and, therefore, light approximately travels in straight lines. Indeed, the branch of optics in which one completely neglects the finiteness of the wavelength is known as geometrical optics (see Chapter 2). (A ray is defined as the path of energy propagation in the limit of $\lambda\to0$.)

Around the second quarter of the nineteenth century, the wave theory seemed to be well established and since it was thought that a wave requires a medium for its propagation, the elastic ether theory was developed. Indeed, in 1823, Fresnel derived the expressions for the reflection and transmission coefficients* by using models for ether vibrations. Poisson (1781–1840), Navier (1785–1836), Cauchy (1789–1857) and many other physicists contributed to the development of the ether theory which also necessitated the development of the theory of elasticity. There were considerable difficulties in the explanation of the models and since we now know that the ether does not exist, we will not go into the details of the various theories.

The nineteenth century also saw the development of electricity and magnetism. In 1820, Oersted (1777–1851) discovered that currents produced magnetic effects. Shortly later, Ampere (1775–1836) found that two parallel wires carrying currents attract each other. Around 1831, Faraday (1791–1867) carried out experiments** which showed that a varying magnetic field induces an electromotive force. Shortly later, Maxwell (1831–1879) generalized Ampere's law by stating that a changing electric field can also set up a magnetic field. He summed up all the laws of electricity and magnetism in the form of equations (which are now referred to as Maxwell's equations***). From these equations, he derived a wave equation and showed the possibility of the existence of electromagnetic waves (see Chapter 19). From the wave equation so derived, he showed that the velocity of the electromagnetic waves can be calculated from experiments in which a certain quantity of electric charge is measured by two different methods. These measurements were carried out by Kohlrausch (1809–1858) and Weber (1804–1891) in 1856 and from their data, Maxwell found that the speed of electromagnetic waves in air should be about 3.1074×10^8 m/sec. He found that this value was very close to the measured value of the speed of light which according to the measurement of Fizeau (1819–1896) in 1849 was known to be 3.14858×10^8 m/sec. The sole fact that the two values were very close to each other led Maxwell to propound his famous electromagnetic theory of light according to which light waves are electromagnetic waves. Associated with a light wave, there would be changing electric and magnetic fields, the changing magnetic field produces a time and space varying electric field and the changing electric field produces a time and space varying magnetic field, this results in the propagation of the electromagnetic wave even in free space. In 1888, Hertz (1857–1894) carried out experiments which could produce and detect electromagnetic waves of frequencies much smaller than that of light. The electromagnetic waves were produced by discharging electrically charged plates through a spark gap. The frequency of the emitted electromagnetic waves depended on the

---

*The derivation of the Fresnel laws on the basis of electromagnetic theory will be discussed in Chapter 20.

**Similar experiments were also carried out by Henry around the same time and the law is also referred to as the Faraday-Henry law.

***The fundamental laws of electricity and magnetism and the Maxwell's equations are discussed in Chapter 19.

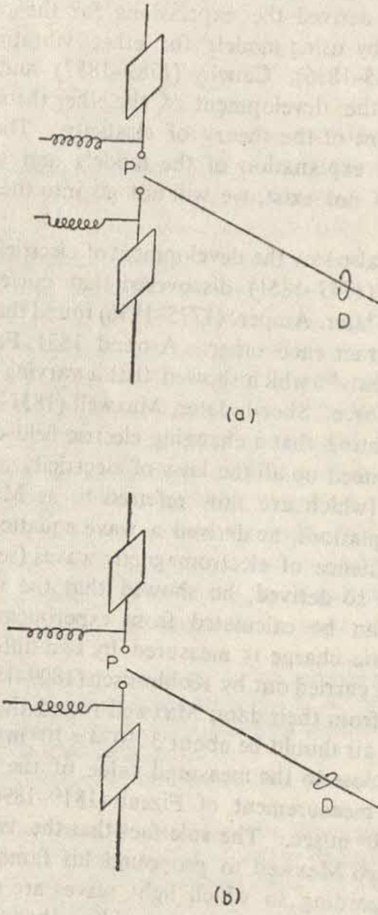


Fig. 1.2 Schematic of Hertz's experiment for generation and detection of electromagnetic waves. Sparks across the gap P produce electromagnetic waves whose frequency depends on the inductance and capacitance of the circuit. The electromagnetic waves can be detected by means of a detector D which is nothing but a short wire bent in the form of a circle with a small gap. A signal is detected if the gap in the detector is parallel to the line joining the knobs of the spark gap P as shown in Fig. 1.2(a); If the gap is at right angles, [see Fig. 1.2(b)], no signal is received.

values of the inductance and capacitance of the circuit. The electromagnetic waves could be detected by means of a detector and it was found that a signal was not received when the detector was placed parallel to the source* (see Fig. 1.2). Hertz also produced standing electromagnetic waves by getting them reflected by a metal sheet (see Figs. 10.3 and 10.4). He could calculate the wavelength of the waves and knowing the frequency, he showed that the speed of the electromagnetic waves was the same as that of light. Using a collimated electromagnetic wave, and getting it reflected by a metal sheet he could demonstrate the laws of reflection. Hertz's experimental results provided dramatic confirmation of Maxwell's electromagnetic theory. In addition, there were so many other experimental results which were quantitatively explained by using Maxwell's theory that around the end of the nineteenth century physicists thought that one finally understood what light really is.

1.4 The Particle Nature of Radiation

In 1887, Hertz discovered that a metal surface irradiated by a light beam would emit electrons if the frequency of the radiation were above a certain critical value. This phenomenon is known as the photoelectric effect and cannot be explained by a theory based on the wave model of light.** In 1905, Einstein (1879–1955) interpreted the photoelectric effect by putting forward his famous photon theory according to which the energy in a light beam of frequency ν was concentrated in corpuscles of energy $h\nu$, where h represents the Planck's constant; these corpuscles are known as photons.*** Einstein's photon theory predicted that if the frequency ν of the incident radiation was greater than the critical frequency ν_c, then the kinetic energy of the emitted electron will be $h(\nu-\nu_c)$ which was later verified by Millikan for visible light, by M. de Broglie for X-rays and by Thibaud and Ellis for γ-rays. Einstein also showed that the photons, in addition to having an energy equal to $h\nu$, should have a momentum $h\nu/c$ which was verified experimentally in 1923 by Compton (1892–1962). This experiment is known as the Compton effect and will be discussed in detail in Sec. 21.3.

---

*This follows from the electromagnetic theory—see Sec. 19.4.

**The detailed experimental findings of the photoelectric effect and the reasons why they cannot be explained on the basis of a wave model will be discussed in Sec. 21.2. It may, however, be mentioned, that when Einstein put forward his photon theory, the detailed results were not available; he was only aware of the fact that below a certain critical frequency of the irradiating beam, there was no electron emission.

***It should be mentioned that in 1900, Planck (1858–1947) had put forward his famous theory of blackbody radiation, the derivation of which presupposed that energy can be absorbed and emitted by an individual resonator only in 'quanta' of magnitude $h\nu$. According to Einstein "...I could nevertheless see to what kind of consequences this law [i.e., Planck's law] of temperature-radiation leads for the photoelectric effect and for other-related phenomena of the transformation of radiation-energy, as well as for the specific heat of (especially) solid bodies. All my attempts, however, to adapt the theoretical foundation of physics to this [new type of] knowledge failed completely. It was as if the ground had been pulled out from under one, with no firm foundation to be seen anywhere, upon which one could have built. That this insecure and contradictory foundation was sufficient to enable a man of Bohr's unique instinct and tact to discover the major laws of the spectral lines and of the electron shells of the atoms together with their significance for chemistry appeared to me like a miracle—and appears to me as a miracle even today. This is the highest form of musicality in the sphere of thought." (Quoted from the autobiographical notes by Einstein in "Albert Einstein: Philosopher Scientist" edited by P.A. Schilpp; Tudor Publishing Co., New York, 1951).

1.5 The Uncertainty Principle

In the previous section, we have mentioned that there are certain experiments which can only be explained on the basis of a corpuscular nature of radiation. The reconciliation of the corpuscular nature with the firmly established wave character of light is through the modern quantum theory; and perhaps the best known consequence of the wave-particle duality is the uncertainty principle of Heisenberg which can be stated as follows: If the x-coordinate of the position of a particle is known to an accuracy Δx, then the x-component of the momentum cannot be determined to an accuracy less than $\Delta p_x \sim \dfrac{h}{\Delta x}$, where h is the Planck's constant. Alternatively, one can say that if Δx and Δp_x represent the accuracies with which the x-coordinate of the position and the x-component of the momentum can be determined, then the following inequality must be satisfied:*

$$\Delta x \, \Delta p_x \gtrsim h \tag{1.5-1}$$

In the present section, we will show how the diffraction of a light beam can be explained on the basis of the corpuscular nature of radiation and the uncertainty principle. Consider a long narrow slit of width b as shown in Fig. 1.3. We assume the light source to be very far away from the slit so that the photons approaching the slit have a momentum only in the y-direction.** Now, according to a classical particle model of radiation

---

*We do not feel the effect of this inequality in our everyday experience because of the smallness of the value of the Planck's constant. For example, for a tiny particle of mass 10^{-5} g, if the position is determined within an accuracy of about 10^{-5} cm, then according to the uncertainty principle, its velocity cannot be determined within an accuracy better than $\Delta v \sim 6\times10^{-17}$ cm/sec. This value is much smaller than the accuracies with which one can determine the velocity of the particle. For a particle of a greater mass, Δv will be even smaller. Indeed, had the value of Planck's constant been much larger, the world would have been totally different. In a beautifully written book, Gamow (1939) has discussed what our world would be like if the Planck's constant had a value of 1 erg-sec so that the effect of the uncertainty principle would be apparent to our sense perception.

**One can always choose the distance between the source and the slit large enough so that p_x can be assumed to have an arbitrarily small value. For example for the source at a distance d, the maximum value of p_x of the photons approaching (as was put forward by Newton), the number of particles reaching the point *P* (which lies in the geometrical shadow) will be extremely small (if not zero); further, if we decrease the width of the slit, the intensity should decrease, which is quite contrary to the experimental results because we know that the beam undergoes diffraction and the intensity at a point like *P* would normally increase if the width of the slit is made smaller. Thus, the classical corpuscular model is quite incapable of explaining the phenomenon of diffraction. However, if we use the uncertainty principle in conjunction with the corpuscular model, the diffraction phenomenon can be explained in the following manner. When a photon passes through the slit, one can say that

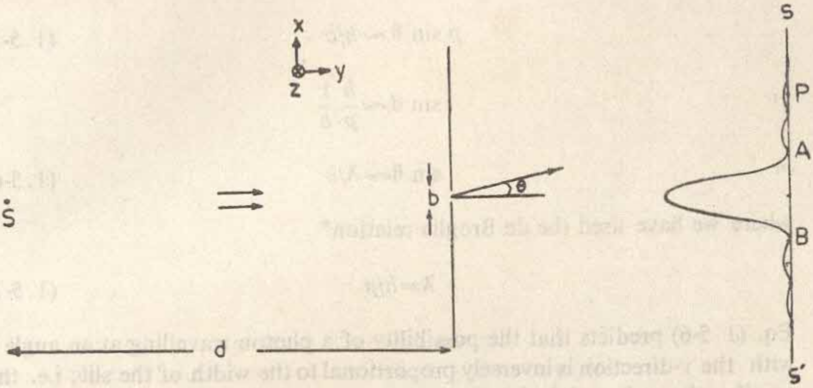


Fig. 1.3 Diffraction of a light beam by a long narrow slit of width *b*

$$\Delta x \sim b \tag{1.5-2}$$

which implies that we can specify the position of the photon to an accuracy *b*. If we now use the uncertainty principle, we would have*

$$\Delta p_x \sim \frac{h}{b} \tag{1.5-3}$$

i.e., just by making the photon pass through a slit of width *b*, the slit imparts a momentum in the x-direction which is $\sim h/b$. Since before the photon entered the slit, $p_x = 0$, we will have

$$|p_x| \sim \Delta p_x \sim h/b \tag{1.5-4}$$

---

the slit will be $\sim p\frac{b}{d} = \frac{h\nu}{c}\,\frac{b}{d}$, which can be made arbitrarily small by choosing a large enough value of *d*.

*Before the photon entered the slit, p_x is known accurately and hence $\Delta p_x = 0$. It would be wrong if we say that by making the photon pass through the slit, $\Delta p_x \Delta x$ is zero, this is because of the fact that $\Delta p_x = 0$ *before* the photon entered the slit. After it has entered the slit, we have no knowledge of p_x.

But $p_x = p \sin\theta$, where θ is the angle that the photon coming out from the slit makes with the y-axis. Thus

$$p \sin\theta \sim h/b \tag{1.5-5}$$

or

$$\sin\theta \sim \frac{h}{p}\frac{1}{b}$$

or

$$\sin\theta \sim \lambda/b \tag{1.5-6}$$

where we have used the de Broglie relation*

$$\lambda = h/p \tag{1.5-7}$$

Eq. (1.5-6) predicts that the possibility of a photon travelling at an angle θ with the y-direction is inversely proportional to the width of the slit; i.e. the smaller the value of b, the greater is the probability of the photon to reach deep inside the geometrical shadow. This is indeed consistent with the diffraction phenomenon (see. Sec. 15.3).

Alternatively, one can use the diffraction phenomenon and the particle nature of radiation to predict the uncertainty principle. We can, therefore, say that the wave-particle duality is a consequence of the uncertainty principle and the uncertainty principle is a consequence of the wave-particle duality.

We conclude this section by quoting Max Born:** "Physicists of today have learnt that not every question about the motion of an electron or a photon can be answered, but only those questions which are compatible with the uncertainty principle."

1.6 The Probabilistic Interpretation of Matter Waves

In the previous section we have seen that if a photon passes through a slit of width b, then the momentum imparted in the x-direction (which is along the width of the slit) is $\sim h/b$. The question arises whether we can predict the trajectory of an individual photon. The answer is no. We cannot say where an individual photon will land up on the screen; we can only

---

*The de Broglie relation follows from the fact that

$$p = \frac{h\nu}{c} = \frac{h}{\lambda}$$

de Broglie had suggested that Eq. (1.5–7) is not only valid for photons but is also valid for all particles like electrons, protons, neutrons, etc. Indeed, the de Broglie relation has been verified by studying the diffraction patterns produced when electrons, neutrons, etc. pass through a single crystal; the patterns can be analysed in a manner similar to the X-ray diffraction patterns (see Chapter 15). In a recent paper, Shull (1969) has studied the Fraunhofer diffraction of neutrons by a single slit and his experimental results agree with the intensity distribution as predicted by Eq. (15.2–10) with λ given by Eq. (1.5–7).

**From Born (1962), Chapter IX.

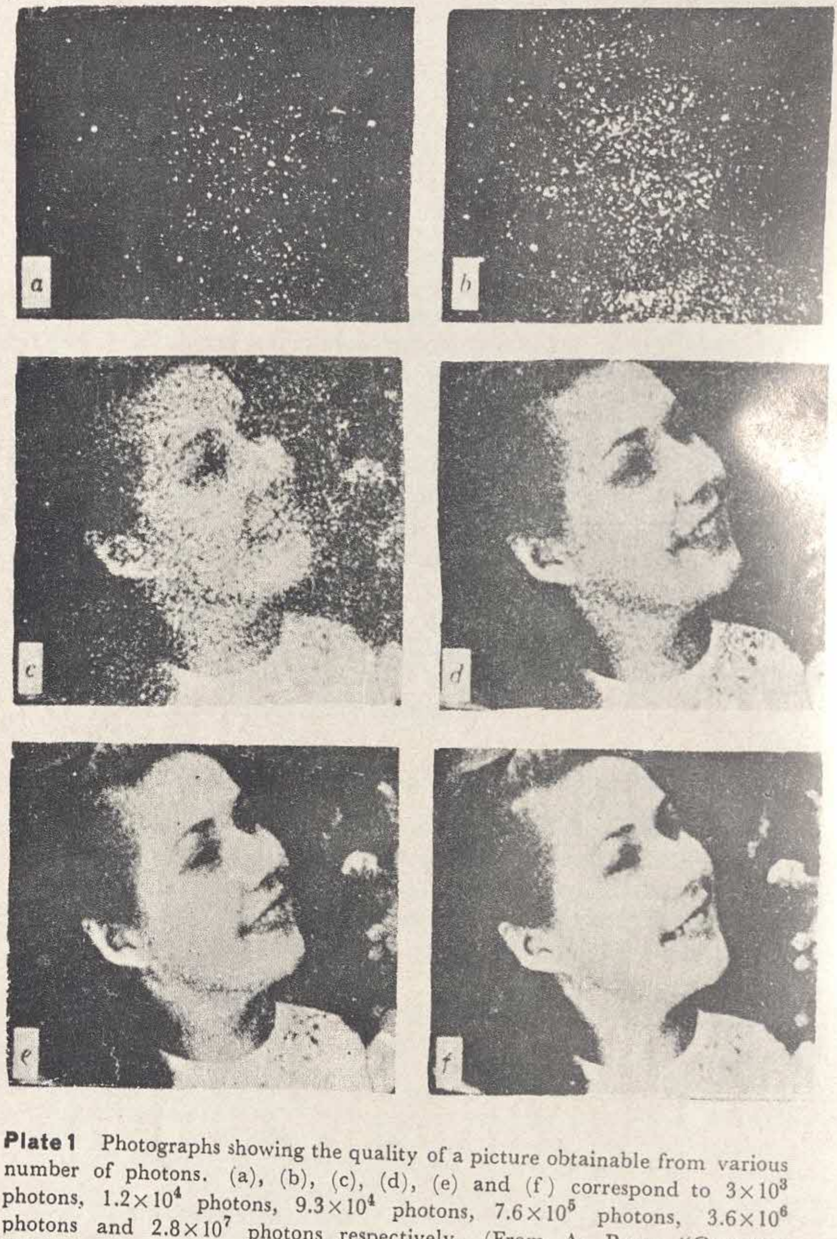


Plate 1 Photographs showing the quality of a picture obtainable from various number of photons. (a), (b), (c), (d), (e) and (f) correspond to 3×10^3 photons, 1.2×10^4 photons, 9.3×10^4 photons, 7.6×10^5 photons, 3.6×10^6 photons and 2.8×10^7 photons respectively. (From A. Rose, "Quantum Effects in Human Vision", *Advances in Biological and Medical Physics*, Vol. V, Academic Press, 1957; used with permission.)

predict the probabilities of arrival of the photon in a certain region of the screen. We may, for example, say that the probability for the arrival of the photon in the region lying between the points *A* and *B* (see Fig. 1.3) is 0.85. This would imply that if the experiment was carried out with a large number of photons, about 85% of them would land up in the region *AB*; but the fate of an individual photon can never be predicted. It should be mentioned that if we place a light detector on the screen, then it will always record one photon or none and never half of a photon. This essentially implies the corpuscular nature of the radiation. However, the probability distribution is roughly the same as predicted by the wave theory and, therefore, if one performs an experiment with a large number of photons (as is indeed the case in most experiments) the intensity distribution recorded on the screen is the same as predicted by the wave theory.

In order to explicitly show that the diffraction is not a many photon phenomenon, Taylor (1909) carried out a beautiful experiment which consisted of a box with a small lamp which casts the shadow of a needle on a photographic plate (see Fig. 1.4). The intensity of the light was so weak that between the slit and the photographic plate, it was almost impossible to find two photons (see Example 1.1). In fact to get a good fringe pattern, Taylor made an exposure lasting for several months. The diffraction pattern, obtained on the photographic plate, was the same as predicted by the wave theory.

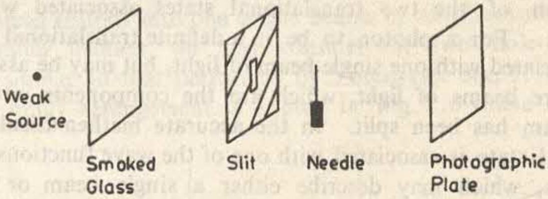


Fig. 1.4 Schematic of the experimental arrangement of Taylor to study the diffraction pattern produced by a weak source. The whole apparatus was placed inside a box.

The corpuscular nature of radiation and the fact that one cannot predict the trajectory of an individual photon can be seen from Plate 1 which consists of series of photographs showing the quality of pictures obtainable from various number of photons. The photograph clearly shows that the picture is built up by the arrival of concentrated packets of energy and the point at which a particular photon will arrive is entirely a matter of chance. The figure also shows that the photograph is featureless when a small number of photons are involved and as the number of photons reaching the photographic plate increases, the intensity distribution becomes the same as would be predicted by the wave theory.

We conclude this section by quoting Feynman:*"...it would be

---

*Quoted from Feynman *et. al.*, (1965).

impossible to predict what would happen. We can only predict the odds! This would mean, if it were true, that physics has given up on the problem of trying to predict exactly what will happen in a definite circumstance. Yes physics *has* given up. *We do not know how to predict what would happen in a given circumstance,* and we believe now that it is impossible—that the only thing that can be predicted is the probability of different events. It must be recognised that this is a retrenchment in our earlier ideal of understanding nature. It may be a backward step, but no one has seen a way to avoid it."

1.7 An Understanding of Interference Experiments

Let us consider the interference experiment involving the Michelson interferometer* in which a light beam is partially reflected by a beam splitter and the resulting beams are made to interfere (see Fig. 1.5). According to Dirac (1958) "...we describe the photon as going partly into each of the two components into which the incident beam is split. The photon is then, as we may say, in a translational state given by the superposition of the two translational states associated with the two components...For a photon to be in a definite translational state it need not be associated with one single beam of light, but may be associated with two or more beams of light, which are the components into which one original beam has been split. In the accurate mathematical theory each translational state is associated with one of the wave functions of ordinary wave optics, which may describe either a single beam or two or more

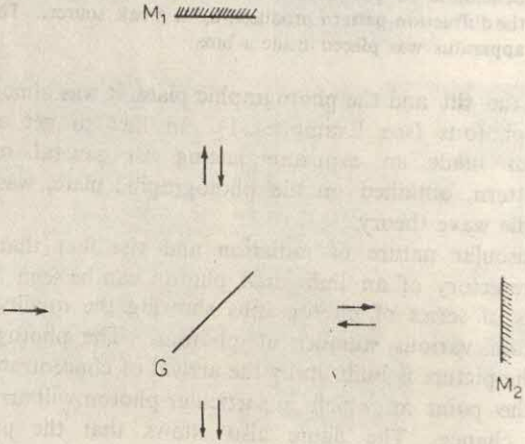


Fig. 1.5 Outline of the Michelson interferometer experiment to obtain an interference pattern. *G* represents a beam splitter, M_1 and M_2 represent plane mirrors.

---

*See Chapter 12.

beams into which one original beam has been split." These translational states can be superposed in a manner similar to the one employed while considering the interference of two beams. Thus, each photon goes partly into each of the two components and interferes only with itself.* If we try to determine the fate of a single photon by measuring the energy in one of the components, then Dirac argues: "The result of such a determination must be either a whole photon or nothing at all. Thus the photon must change suddenly from being partly in one beam and partly in the other to be entirely in one of the beams. This sudden change is due to the disturbance in the translational state of the photon which the observation necessarily makes. It is impossible to predict in which of the two beams the photon will be found. Only the probability of either result can be calculated....Our description of the photon allows us to infer that, *after* such an energy measurement, it would not be possible to bring about any interference effects between the two components. So long as the photon is partly in one beam and partly in the other, interference can occur when the two beams are superposed, but this possibility disappears when the photon is forced entirely into one of the beams by an observation."

In a similar manner, we may consider the two hole interference experiment similar to that performed by Young (see Sec. 11.4 and 11.5). The experimental arrangement is shown in Fig. 1.6 where a weak light

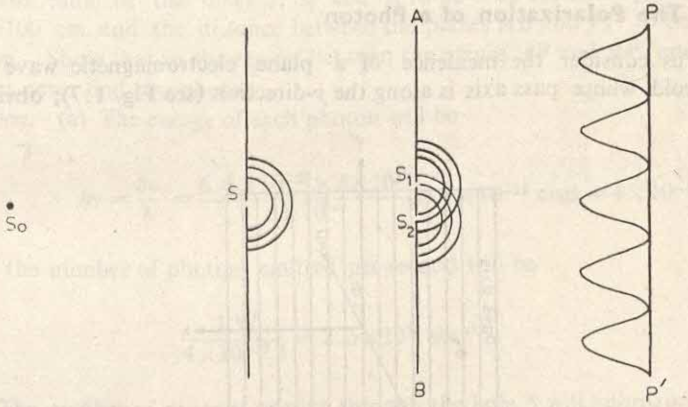


Fig. 1.6 Young' double hole experimental arrangement for obtaining the interference pattern. S_0 represents a point source.

source S_0 illuminates the hole S and the light beams emerging from the holes S_1 and S_2 produce the interference pattern on the screen PP'. The intensity is assumed to be so weak that in the region between the planes

---

*Dirac has further stated that "interference between two different photons never occurs." This statement has lead to a lot of controversy particularly in the light of recent experiments on the interference of two laser beams. For a detailed interpretation one may look up Mandel and Wolf (1965).

AB and PP' there is almost never more than one photon (see Example 1.1 and Sec. 1.8). Individual photons are also counted by a detector on the screen PP' and one finds that the intensity distribution has a $\cos^2$ pattern similar to that shown in Figs. 11.8 (a) and (b). The corpuscular nature of the radiation is evident from its detection in the form of single photons and never a fraction of a photon. The appearance of the interference pattern is because of the fact that a photon interferes with itself. The quantum theory tells us that a photon partially passes through the hole S_1 and partially through S_2. This is not the splitting of the photon into two halves but only implies that if we wish to find out through which hole the photon passed through,* then half the time it will be found to have passed through the hole S_1 and half the time through S_2. As in the case of the Michelson interferometer, the wave function associated with the state of the photon is a superposition of two wavefunctions, one corresponding to the waves emanating from hole S_1 and the other emanating from hole S_2. The superposed wavefunction will give rise to an intensity distribution similar to that obtained by considering the superposition of two waves (see Chapter 11).

Thus we may say that the photons would arrive as packets of energy but the probability distribution (on the screen) will be proportional to the intensity distribution predicted by using a wave model.

1.8 The Polarization of a Photon

Let us consider the incidence of a plane electromagnetic wave on a polaroid whose pass axis is along the y-direction (see Fig. 1.7); obviously,

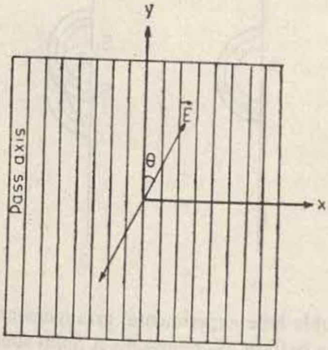


Fig. 1.7 The incidence on a polaroid of linearly polarized light beam whose electric vector makes an angle θ with the y-axis; the pass-axis of the polaroid is along the y-axis.

---

*It may be mentioned that if we had a method of determining the hole through which the photons passed through then the interference pattern on the screen will be washed out. This is a consequence of the fact that a measurement always disturbs the system. For further details, one may look up Feynman *et al.* (1964).

the electric vector of the transmitted wave would be along the y-direction (see Chapter 18). Thus, if the electric-vector associated with the incident wave oscillates along the x-axis, then, the wave will be absorbed by the polaroid; on the other hand, if the electric-vector oscillates along the y-axis, it will just pass through the polaroid. Further, if the electric-vector makes an angle θ with the pass axis, then the intensity of the transmitted beam will be $I_0 \cos^2 \theta$, where I_0 represents the intensity of the incident beam (this is known as Malus' law which will be discussed in Sec. 18.3).

In the photon theory also, one can associate a certain state of polarization with every photon. One can argue that if the electric-vector associated with the photon is along the y-(or the x-axis), then the photon will pass through (or get absorbed) by the polaroid. The question now arises as to what would happen to a single photon if the **E** makes an angle θ with the pass axis. The answer to this question is that the probability for the photon to pass through the polaroid is $\cos^2 \theta$ and if the experiment is conducted with N photons (and if N is very large) then about $N \cos^2 \theta$ photons will pass through; one cannot predict the fate of an individual photon.

Example 1.1 Let a source (with $\lambda = 5 \times 10^{-5}$ cm) of power 1 W be used in the experimental arrangement shown in Fig. 1.6. (a). Calculate the number of photons that are being emitted by the source per second. (b) Assume that the radii of the holes S, S_1 and S_2 to be 0.02 cm and $S_0S = SS_1 = SS_2 = 100$ cm and the distance between the planes AB and PP' to be also 100 cm. Show that in the region between the planes AB and PP' one can almost never find two photons.

Solution. (a) The energy of each photon will be

$$h\nu = \frac{hc}{\lambda} = \frac{6.6 \times 10^{-27} \times 3 \times 10^{10}}{5 \times 10^{-5}} \simeq 4 \times 10^{-12} \text{ ergs} = 4 \times 10^{-19} \text{ J}$$

Thus the number of photons emitted per second will be

$$\frac{1 \text{ W}}{4 \times 10^{-19} \text{ J}} = 2.5 \times 10^{18} \text{ sec}^{-1}$$

(b) The number of photons passing through the hole S will approximately be

$$\frac{2.5 \times 10^{18} \times \pi \times (0.02)^2}{4 \times \pi \times (100)^2} \simeq 2.4 \times 10^{10} \text{ per second.}$$

Similarly, the number of photons passing either S_1 or S_2 will approximately be

$$\frac{0.6 \times 10^{10} \times 2 \times \pi \times (0.02)^2}{4 \times \pi \times (100)^2} \simeq 120 \text{ per second.}$$

The distance between the planes AB and PP' is 100 cm which will be traversed by the photon in a time$\sim 3\times10^{-9}$ sec. Thus (approximately) every hundredth of a second a photon enters the region and the space is traversed much before the second photon enters.* Therefore, in the region between AB and PP' one will (almost) never find two photons.

PROBLEMS

1. The position of a particle (of mass 10^{-3} g) can be determined to an accuracy $\sim10^{-5}$ cm. Calculate the time for the particle to move a detectable distance. (Ans: about 500,000 years !)
2. Calculate the de Broglie wavelength for a particle of mass 10^{-4} g with speed 1 cm/sec; compare with that of the de Broglie wavelength of 1 eV electrons and 1 eV neutron. Assume $m_e \simeq 9.11\times10^{-28}$ g and $m_n \simeq 1.67\times10^{-24}$ g, where m_e and m_n represent the masses of the electron and of the neutron respectively.
3. Thermal neutrons corresponding to $T \simeq 300$°K can be used for studying crystal structures. Why? [Thermal neutrons have kinetic energies k_BT, where k_B ($\simeq 1.380\times10^{-23}$ J/°K) represents the Boltzmann's constant and T the absolute temperature].
4. (a) A particle of mass 10^{-4} g and speed 1 cm/sec passes through a slit of width 0.01 cm. What will be the uncertainty in the diffraction angle?
   (b) Compare this with the result that would be obtained for a 10 eV electron and for a photon corresponding to $\lambda = 6\times10^{-5}$ cm. (Ans: (a) $\sim6\times10^{-21}$ radians).
5. A 1 W laser beam (of diameter 2 cm) falls on two circular holes each of diameter 0.05 cm as shown in Fig. 1.9. Calculate the average number of photons that will be found between the planes AB and PQ. Assume $\lambda \simeq 6\times10^{-5}$ cm, and the distance between the planes AB and PQ to be 30 cm. (Ans: $\sim4\times10^{6}$ photons).

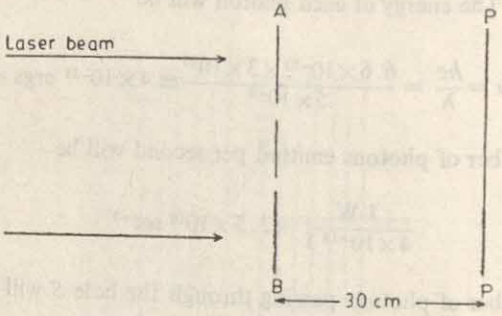


Fig. 1.9 For Problem 6

6. The uncertainty principle can be used for the determination of the size of the atom. Let us consider the simplest atom (namely hydrogen) which consists of an electron and a proton. If we assume that the electron is confined to a region $\sim a$, then uncertainty principle tells us that $p \sim \Delta p \sim \frac{\hbar}{a}$, where $\hbar \equiv \frac{h}{2\pi}$ (we have used $\hbar$

*It is somewhat similar to the case when, on an average, 10 persons pass through a room in one year and the time that each person takes to cross the room is $\sim$1 sec, thus, it will be highly improbable to have two persons simultaneously in the room.

instead of h to get the correct result!). Thus, kinetic energy $\sim\frac{\hbar^2}{2ma^2}$ and potential energy $\sim -e^2/a$, where e is the charge of the electron. (a) Minimize the total energy and show that $a=\frac{\hbar^2}{me^2}\simeq\frac{1}{2}$ Å, which represents the Bohr radius; this represents the size of the hydrogen atom. (Since the proton is much heavier than the electron, we have assumed it to be at rest). (b) Show that the corresponding energy is $-\frac{me^4}{2\hbar^2}\simeq -13\cdot 6$ eV which is the ionization energy of the hydrogen atom.

7. Use the analysis of the above problem to calculate the ground state energy of helium atom; compare it with the experimental value of 5.8 eV. (Hint: A helium atom consists of two electrons for which the total energy will be given by

$$E\simeq\frac{\hbar^2}{2m}\left(\frac{1}{a_1^2}+\frac{1}{a_2^2}\right)-2e^2\left(\frac{1}{a_1}+\frac{1}{a_2}\right)+\frac{e^2}{a_1+a_2}$$

Use the conditions $\frac{\partial E}{\partial a_1}=0=\frac{\partial E}{\partial a_2}$ to determine a_1, a_2 and E).

8. Using the uncertainty principle show that the ground state energy of a linear harmonic oscillator is $\sim\hbar\omega$. (Hint: The total energy of a linear harmonic oscillator is $\frac{p^2}{2m}+\frac{1}{2}m\omega^2x^2$).,

9. When an atom jumps from an energy state E_2 to an energy state $E_1(E_2>E_1)$ a photon of frequency $\nu=(E_2-E_1)/h$ is emitted. This emission is essentially a pulse of duration $\sim 10^{-9}$ sec, this duration is usually denoted by τ. This leads to a frequency width $\Delta\nu$ and as will be shown in Chapter 14,

$$\tau\Delta\nu\sim 1$$

Multiplying both sides by h, we get the time-energy uncertainty principle:

$$\Delta t\,\Delta E \gtrsim h$$

which can be interpreted as follows: If Δt represents the uncertainty in the time at which a time dependent process takes place, then the uncertainty ΔE in the energy of this process will be $\gtrsim h/\Delta t$. Assuming $\Delta t\sim 10^{-9}$ sec, calculate the energy width of the excited state.

REFERENCES AND SUGGESTED READING

BAKER, A. (1970), *Modern Physics and Antiphysics*, Addison-Wesley Publishing Co., Reading, Mass (USA).

BARTON, A. W. (1939), *A Textbook on Light*, Longmans Green & Co., London (England).

BORN, M. (1969), *Atomic Physics*, eighth edition, Blackie & Son Ltd., London (England).

DE BROGLIE, L. (1937), *Matter and Light: The New Physics*, Dover Publications, New York (USA).

DIRAC, P. A. M. (1963), The evolution of the physicists picture of nature, in *Project Physics Reader 5—Models of the Atom*, Harvard Project Physics.

FEINBERG, G. (1968), Light, *Scient. Amer.*, **219**, September, issue, p. 50.

FEYNMAN, R. P., LEIGHTON, R. B. and SANDS, M. (1965), *The Feynman Lectures on Physics*, Vol. III Addison-Wesley Publishing Co., Reading, Mass. (USA).

GAMOW, G. (1940), *Mr. Tompkins in Wonderland*, Cambridge University Press, London (England).

HEISENBERG, W. (1949), *The Physical Principles of the Quantum Theory*, Dover Publications, New York (USA).

JOYCE, W. B. and JOYCE, A. (1976), Descartes, Newton and Snell's law, *J. Opt. Soc. Amer.* 66, 1.

MANDEL, L. and WOLF, E. (1965), Coherence properties of optical field, *Rev. Mod. Phys.*, 37, 231.

OPPENHEIMER, J. R. (1969), *The Flying Trapeze: Three Crises for Physicists*, Harper and Row, New York (USA).

ROSE, A. (1957), Quantum effects in human vision, *Advances in Biological and Medical Physics*, Vol. V., Academic Press, New York (USA).

SCHILPP, P. A. (1951), *Albert Einstein: Philosopher Scientist*, (Ed. P. A. Schilpp), Tudor Publishing Co., New York (USA).

SHULL, C. G. (1969), Single slit diffraction of neutrons, *Phys. Rev.*, 179, 752.

SIMMONS, J. W. and GUTTMANN, M. J. (1970), *States, Waves and Photons: A Modern Introduction to Light*, Addison-Wesley Publishing Co., Reading, Mass. (USA).

WICHMANN. E. H. (1971), *Quantum Physics: Berkeley Physics Course*, Vol. 4, McGraw-Hill Book Co., New York (USA).

PART I

Geometrical Optics

S.C.E.R.T., West Bengal
Date 15.3.91
Acc. No. 4958

S.C.E.R.T. West Bengal
Library
25/3, B. C. Road
Calcutta

2 Fermat's Principle and its Applications

> *With the genius, nature remains in eternal union:*
> *What the one promises, the other certainly redeems.*
>
> SCHILLER

2.1 Introduction

The study of the propagation of light in the realm of geometrical optics employs the concept of rays. To understand what a ray is, consider a circular aperture in front of a point source *P* as shown in Fig. 2.1. Wher

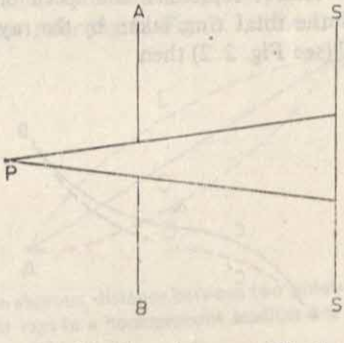


Fig. 2.1 The light emitted by the point source *P* is allowed to pass through a circular hole and if the diameter of the hole is very large compared to the wavelength of light then the light patch on the screen *SS′* has well defined boundaries.

the diameter of the aperture is quite large (~ 1 cm), then on the screen *SS′*, one can see a patch of light with well defined boundaries. When we start decreasing the size of the aperture, then at first the size of the patch starts decreasing, but when the size of the aperture becomes very small ($\lesssim 0.1$ mm) then the pattern obtained on SS' does not have well defined boundaries. This phenomenon is known as diffraction* and is a direct consequence of the finiteness of the wavelength (which is denoted by λ). It will be shown in Chapter 15 that the diffraction effects become smaller with decrease in wavelength and indeed in the limit of $\lambda \to 0$, the diffraction effects will be absent, and even for extremely small diameters of the aperture, we will obtain a well defined shadow on the screen SS'; and, therefore, in the zero wavelength limit one can obtain an infinitesimally thin pencil of light; this is called a *ray*. Thus, a ray defines the path of propagation of the energy in the limit of the wavelength going to zero. Since light has a wavelength of the order of 10^{-5} cm, which is small compared to the dimensions of normal optical instruments like lenses, mirrors, etc. one can, in many applications, neglect the finiteness of the wavelength. The field of optics under such an approximation (i.e. the neglect of the finiteness of the wavelength) is called geometrical optics.

The field of geometrical optics can be studied by using Fermat's principle which determines the path of the rays. According to this principle 'the ray will correspond to that path for which the time taken is an extremum in comparison to nearby paths', i.e. it is either a minimum or a maximum or stationary.** Let $n(x, y, z)$ represent the position dependent refractive index. Then

$$\frac{ds}{c/n} = \frac{n\,ds}{c}$$

will represent the time taken to traverse the geometric path ds in a medium of refractive index n. Here c represents the speed of light in free space. Thus, if τ represents the total time taken by the ray to traverse the path AB along the curve C (see Fig. 2.2) then

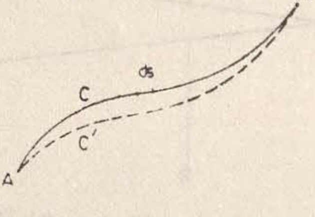


Fig. 2.2 If the path ACB represents the ray path then the time taken in traversing the path ACB will be an extremum in comparison to any nearby path $AC'B$; i.e. it will either correspond to a minimum or a maximum or stationary value of the optical path length.

---

*The phenomenon of diffraction is discussed in detail in Chapters 15 and 16.

**Fermat's principle can be derived from Maxwell's equations; see, for example, Born and Wolf (1975).

$$\tau = \frac{1}{c}\int_{\substack{A\to B\\ C}} n\,ds \tag{2.1-1}$$

where the symbol $\underset{C}{A\to B}$ below the integral represents the fact that the integration is from the point A to B through the curve C. Let τ' be the time taken along the nearby path $AC'B$ (shown as the dashed curve in Fig. 2.2), and if ACB indeed represents the path of a ray, then τ will be either less than, greater than or equal to τ' for *all* nearby paths like $AC'B$. Thus according to Fermat's principle, out of the many paths connecting two points, the light ray would follow that path for which the time taken is an extremum. Since c is a constant, one can alternatively define a ray as the path for which

$$\int_{\substack{A\to B\\ C}} n\,ds \tag{2.1-2}$$

is an extremum; the above integral represents the optical path from A to B along C; i.e., the ray would follow the path for which

$$\delta \int_{A\to B} n\,ds = 0 \tag{2.1-3}$$

From the above principle one can immediately see that in a homogeneous medium (i.e. in a medium whose refractive index is constant everywhere), the rays will be straight lines because a straight line will correspond to a minimum value of the optical path connecting two points in the medium. Thus referring to Fig. 2.3, if A and B are two points in

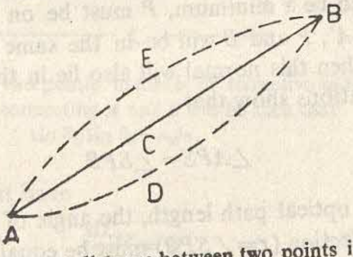


Fig. 2.3 Since the shortest distance between two points is along a straight line, light rays in a homogeneous medium are straight lines.

a homogeneons medium, then the ray path will be along the straight line ACB because any nearby path like ADB or AEB will correspond to a longer time.

2.2 Laws of Reflection and Refraction from Fermat's Principle

We will now obtain the laws of reflection and refraction from Fermat's principle. Consider a plane mirror MN as shown in Fig. 2.4. To obtain the laws of reflection, we have to determine that path from A to B (via the mirror) which has the minimum optical path length. Since the path would lie completely in a homogeneous medium, we need to minimize only the path length. Thus we have to find that path APB for which $AP+PB$ is a minimum. To find the position of P on the mirror, we drop a perpendicular from A on the mirror and let A' be a point on the perpendicular such that $AR=RA'$; thus $AP=PA'$ and $AQ=A'Q$ where AQB is another

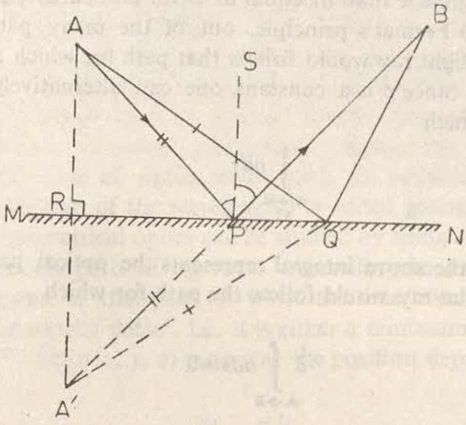


Fig. 2.4 The shortest path connecting the two points A and B via the mirror is along APB where the point P is such that AP, PS and PB are in the same plane and $\angle APS=\angle SPB$; PS being the normal to the plane of the mirror. The straight line path AB is also a ray.

path adjacent to APB. Thus we have to minimize the length $A'PB$. Clearly, for $A'PB$ to be a minimum, P must be on the straight line $A'B$. Thus the points A, A', P and B will be in the same plane and if we draw a normal PS at P then this normal will also lie in the same plane. Simple geometric considerations show that

$$\angle APS=\angle SPB$$

Thus for minimum optical path length, the angle of incidence ($i=\angle APS$) and the angle of reflection ($r=\angle SPB$) must be equal and the incident ray, the reflected ray and the normal to the surface at the point of incidence on the mirror must be in the same plane. These form the laws of reflection. It should be pointed that, in the presence of the mirror there will be two ray paths which will connect the points A and B; the two paths will be AB and APB. Fermat's principle tells us that whenever the optical path length is an extremum, we will have a ray, and thus, in general, there may be more than one ray path connecting two points.

To obtain the laws of refraction, let PQ be a surface separating two media of refractive indices n_1 and n_2 as shown in Fig. 2.5. Let a ray starting from the point A, intersect the interface at R and proceed to B along RB. Clearly, for minimum optical path length, the incident ray, the refracted ray and the normal to the interface must all lie in the same plane. To determine that point R for which the optical path length from A to B is a minimum, we drop perpendiculars AM and BN from A and B respectively on the interface PQ. Let $AM=h_1$, $BN=h_2$ and $MR=x$. Then since A and B are fixed, $RN=L-x$, where $MN=L$ is a fixed quantity. The optical path length from A to B, by definition, is

$$L_{op}=n_1AR+n_2RB=n_1\sqrt{x^2+h_1^2}+n_2\sqrt{(L-x)^2+h_2^2} \qquad (2.2\text{-}1)$$

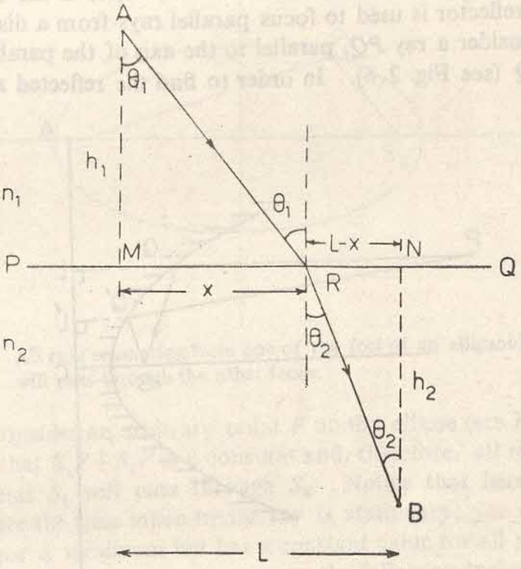


Fig. 2.5 A and B are two points in media of refractive indices n_1 and n_2. The ray path connecting A and B will be such that $\sin\theta_1/\sin\theta_2=n_2/n_1$.

To minimize this, we must have

$$\frac{dL_{op}}{dx}=0$$

i.e.,

$$\frac{n_1x}{\sqrt{x^2+h^2}}-\frac{n_2(L-x)}{\sqrt{(L-x)^2+h}}=0 \qquad (2.2\text{-}2)$$

Further, as can be seen from Fig. 2.5

$$\sin\theta_1=x/\sqrt{x^2+h_1^2} \qquad (2.2\text{-}3)$$

and

$$\sin\theta_2=(L-x)/\sqrt{(L-x)^2+h_2^2} \qquad (2.2\text{-}4)$$

Thus Eq. (2.2-2) becomes

$$n_1 \sin \theta_1 = n_2 \sin \theta_2 \tag{2.2-5}$$

which is the Snell's law of refraction.

The laws of reflection and refraction form the basic laws for tracing light rays through simple optical systems, like a system of lenses and mirrors, etc.

Example 2.1 Consider a set of rays, parallel to the axis, and incident on a paraboloidal reflector. Show, by using Fermat's principle, that all the rays will pass through the focus of the paraboloid. (A paraboloid is obtained by rotating a parabola about its axis.) This is the reason why a paraboloidal reflector is used to focus parallel rays from a distant source.

Solution. Consider a ray *PQ*, parallel to the axis of the parabola, incident at the point *Q* (see Fig. 2.6). In order to find the reflected ray, one has to draw a normal at the point *Q* and then draw the reflected ray. It can be shown from geometrical considerations that the reflected ray *QS* will always pass through the focus *S*. However, this procedure will be quite cumbersome and as we will show below, the use of the Fermat's principle leads us to the desired result immediately.

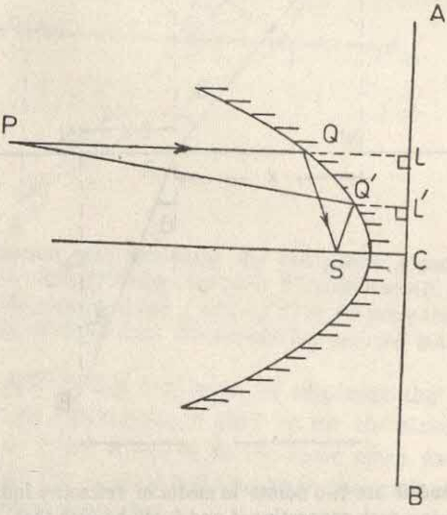


Fig. 2.6 All rays parallel to the axis of a paraboloidal reflector pass through the focus after reflection. The line *ACB* is the directrix.

In order to use Fermat's principle we try to find out the ray connecting the focus *S* and an arbitrary point *P* (see Fig. 2.6). Let the ray path be *PQ′S*. According to Fermat's principle the ray path will correspond to a minimum value of $PQ'+Q'S$. From the point Q' we drop a perpendicular, $Q'L'$, on the directrix *AB*. From the definition of the parabola it follows that $Q'L'=Q'S$. Thus

$$PQ'+Q'S=PQ'+Q'L'$$

Let L be the foot of the perpendicular drawn from the point P on AB. Then, for $PQ'+Q'L'$ to be a minimum, the point Q should lie on the straight line PL, and thus the actual ray which connects the point P and S will be $PQ+QS$ where PQ is parallel to the axis. Therefore, all rays parallel to the axis will pass through S and conversely all rays emanating from the point S, will become parallel to the axis after suffering a reflection.

Example 2.2 Consider an elliptical reflector whose foci are the points S and S_2 (see Fig. 2.7). Show that all rays emanating from the point S_1 wil pass through the point S_2 after undergoing a reflection.

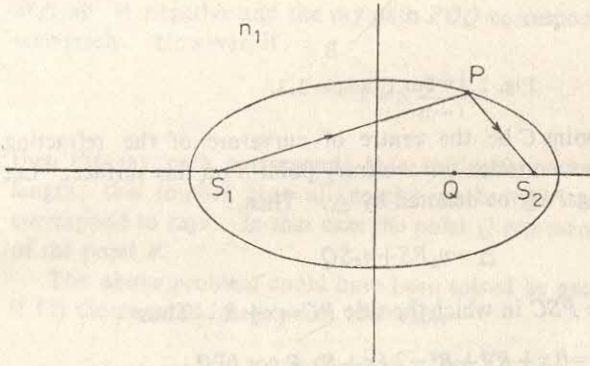


Fig. 2.7 All rays emanating from one of the foci of an ellipsoidal reflector will pass through the other focus.

Solution. Consider an arbitrary point P on the ellipse (see Fig. 2.7). It is well known that S_1P+S_2P is a constant and, therefore, all rays emanating from the point S_1 will pass through S_2. Notice that here we have an example where the time taken by the ray is stationary; i.e. it is neither a maximum nor a minimum but has a constant value for all points lying on the mirror. As a corollary, we may note the following two points:

(i) Excepting the rays along the axis, no other ray (emanating from either of the foci) will pass through an arbitrary point Q which lies on the axis.

(ii) The above considerations will remain valid even for an ellipsoid of revolution obtained by rotating the ellipse about its major axis.

Example 2.3 Consider a refracting surface of radius of curvature R as shown in Fig. 2.8. We consider two points P and Q such that $PO=OQ=x$ (say). Assume that the centre of curvature C lies on the line joining P and Q. Show that the straight line path POQ corresponds to an extremum, and this extremum will be a minimum when $x<R\frac{n+1}{n-1}$; a maximum when $x>R\frac{n+1}{n-1}$ and stationary when $x=R\frac{n+1}{n-1}$. The stationary value will correspond to the paraxial image of the point P. Here $n=n_2/n_1$ where n_1 and n_2 are the refractive indices of the media on the left and on the right of the refracting surface.

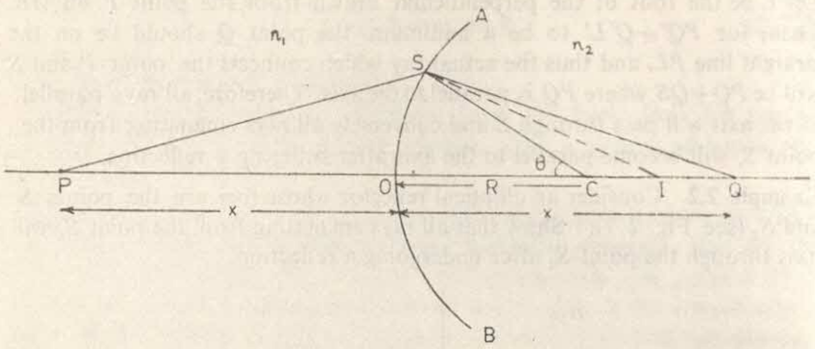


Fig. 2.8 For Example 2.3.

Solution. Let the point C be the centre of curvature of the refracting surface AOB. Let us consider an arbitrary point S on this surface. Let the optical path along PSQ be denoted by $\triangle$. Then,

$$\triangle = n_1 PS + n_2 SQ$$

Consider the triangle PSC in which the side $PC = x+R$. Thus

$$PS = [(x+R)^2 + R^2 - 2\,(x+R)\,R\cos\theta]^{1/2}$$

where θ has been defined in the figure. For small values of θ

$$PS \simeq \left[x^2 + 2xR + R^2 + R^2 - 2\,(xR + R^2)\left(1 - \frac{\theta^2}{2}\right)\right]^{1/2}$$

$$\simeq x\left[1 + \frac{(x+R)\,R\theta^2}{x^2}\right]^{1/2} \simeq x + \frac{(x+R)\,R\theta^2}{2x}$$

Similarly,

$$SQ \simeq x - \frac{(x-R)\,R\,\theta^2}{2x}$$

Thus

$$\triangle \simeq n_1 x + n_1 \frac{(x+R)\,R\theta^2}{2x} + n_2 x - n_2 \frac{(x-R)\,R\theta^2}{2x}$$

Therefore,

$$\frac{d\triangle}{d\theta} \simeq n_1 \frac{(x+R)\,R}{x}\theta - n_2 \frac{(x-R)\,R}{x}\theta$$

Clearly $\frac{d\triangle}{d\theta} = 0$ when $\theta = 0$. Thus the path POQ always corresponds to an extremum and, therefore, it is always a ray path.*

In order to determine whether the extremum corresponds to a maximum or a minimum we calculate $d^2\triangle/d\theta^2$ at $\theta = 0$:

---

*This could have followed directly from the fact that the ray PO is incident normally on the surface and the refracted ray would go undeviated.

$$\left.\frac{d^2\Delta}{d\theta^2}\right|_{\theta=0}=\frac{n_1\,(x+R)\,R}{x}-\frac{n_2\,(x-R)\,R}{x}$$
$$=n_1\frac{R}{x}\,[R\,(n+1)-x\,(n-1)]$$

where $n=n_2/n_1$ represents the relative refractive index, which we assume to be greater than 1. Clearly, if

$$x>R\frac{n+1}{n-1},$$

$d^2\Delta/d\theta^2$ is negative and the ray path POQ corresponds to a minimum and conversely. However, if

$$x=R\frac{n+1}{n-1},$$

then the ray path corresponds to a stationary value of the optical path length; this implies that all nearby paths (for small values of θ) also correspond to rays. In that case the point Q represents the paraxial image of the point P.

The above problem could have been solved by geometry by noting that if I is the paraxial image point of P then

$$n_1PO+n_2OI=n_1PS+n_2SI$$

(The above equation defines the paraxial image point.) Thus

$$n_1PO+n_2OQ=n_1PS+n_2\,(SI-OI+OQ)$$
$$=n_1PS+n_2\,(SI+IQ)>n_1PS+n_2SQ$$

Similarly, if Q lies on the left of I then n_1PO+n_2OQ would have been less than n_1PS+n_2SQ. Notice that in this proof we do not require OP to be equal to OQ.

2.3 Application of Fermat's Principle to Inhomogeneous Media

On a hot day, the air near the ground has a higher temperature than the air which is much above the surface. Since the density of air is lesser at higher temperatures, a refractive index gradient is set up; i.e. the refractive index increases as we go up from the ground. In such a case the rays are curved and one obtains the formation of a mirage; the rays reaching the eye appear to be coming from a point below the ground.

From the temperature distribution we can obtain the refractive index variation which can be approximately assumed to be of the form*

$$n^2\,(x)=n_0^2+n_p^2\,(1-e^{-\alpha x}),\ x>0 \qquad (2.3\text{-}1)$$

---

*The discussion is based on a paper by Khular, Thyagarajan and Ghatak (1977); see also Sodha, Aggarwal and Kaw (1967) and Fraser and Mach (1976).

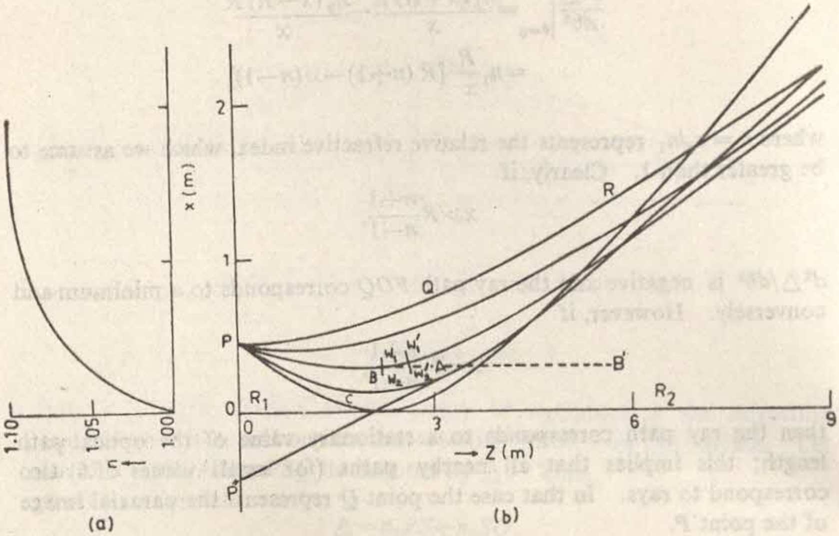


Fig. 2.9 (a) Refractive index variation for an inhomogeneous medium characterized by Eqs. (2.3-1) and (2.3-2). (b) Corresponding ray paths for an object point at a height of $1/\alpha$ ($\simeq$0.43m). [Adapted from Khular, Thyagarajan and Ghatak (1977).]

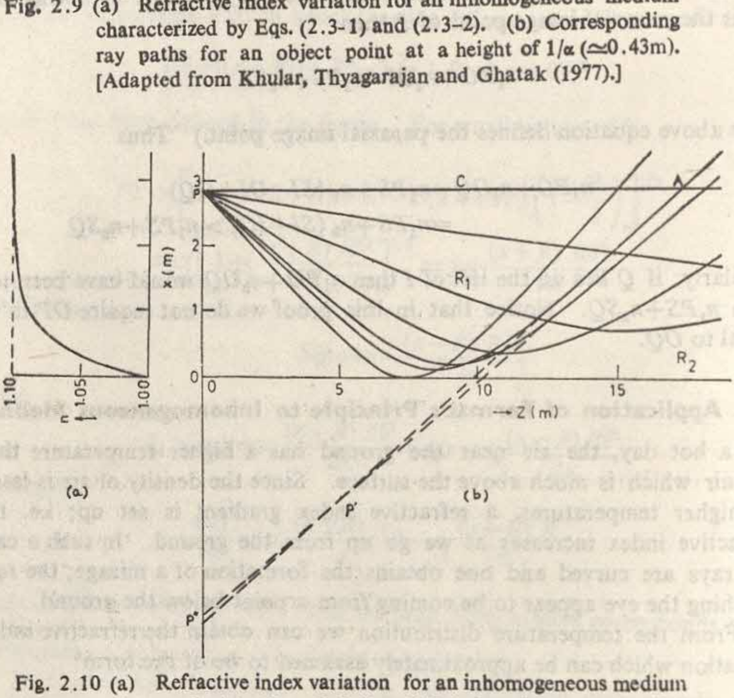


Fig. 2.10 (a) Refractive index variation for an inhomogeneous medium characterized by Eqs. (2.3-1) and (2.3-2). (b) Corresponding ray paths for an object point at a height of about 2.8 m., [Adapted from Khular, Thyagarajan and Ghatak (1977)].

where n_0, n_p and α are constants and x is the height above the ground. The refractive index at $x=0$ is n_0 and for large values of x it approaches the value $(n_0^2+n_p^2)^{1/2}$. The ray paths are obtained by solving the ray equation, which will be discussed later in this section. Typical ray paths (emanating from a point P) are shown in Figs. 2.9 and 2.10. The figures correspond to

$$n_0=1.000233;\ n_p=0.45836;\ \text{and}\ \alpha=2.303\ m^{-1} \qquad (2.3\text{-}2)$$

The refractive index variation is shown in Figs. 2.9 (a) and 2.10 (a). It can be seen from Fig. 2.9 (b) that when A and P are close to the ground, the only ray path connecting the points P and A will be along the curve PCA and a ray emanating horizontally from the point P will propagate in the upward direction as PQR as shown in Fig. 2.9 (b). Thus, under such a condition, the eye at A will see the image of P at P'' and will *not see* the object directly at P. However, if the points P and A are much above the ground [see Fig. 2.10 (b)] then the eye will see the object directly (because of the rays like PCA); it will also receive rays which will appear to emanate from points like P' and P'', this will give rise to mirage. In both cases, however, there is a well defined region R_1 where only the object is visible and the virtual image is not seen. Another interesting region is R_2, where none of the rays (emanating from the point P) reaches; thus, an eye in this region will neither see the object nor its image. This region is hence called the shadow region.

The formation of a mirage occurs due to a refractive index gradient in the air above a hot surface. However, above cold sea water, the air near the surface is colder than the air above it and hence there is an opposite temperature gradient. A suitable refractive index variation for such a case can be written as

$$n^2 = n_0^2 + n_p^2\, e^{-\alpha x} \qquad (2.3\text{-}3)$$

If we solve the ray equation for this profile (see Problem 2.6), the ray paths obtained will be as shown in Fig. 2.11. In this case, an eye at a point A will receive rays which will appear to emanate from points like P' and P'' (which are above P). Such a phenomenon, in which the object appears to be above its actual position is known as 'looming'; it is commonly observed over cold sea waters. Further, since no other rays emanating from P reach A, the object will not be visible directly.

In order to obtain the precise ray path in such media, we will derive the ray equation for an inhomogeneous medium. We will restrict ourselves to the special case when the refractive index changes continuously along one direction only; this direction we assume to be along the x-axis. This medium can be thought of as a limiting case of a medium consisting of a continuous set of thin slices of media of different refractive indices. When the thickness of each slice goes to zero and at the same time the number of slices increases indefinitely, then one obtains a medium in which the refractive index changes continuously along the x-direction. At each

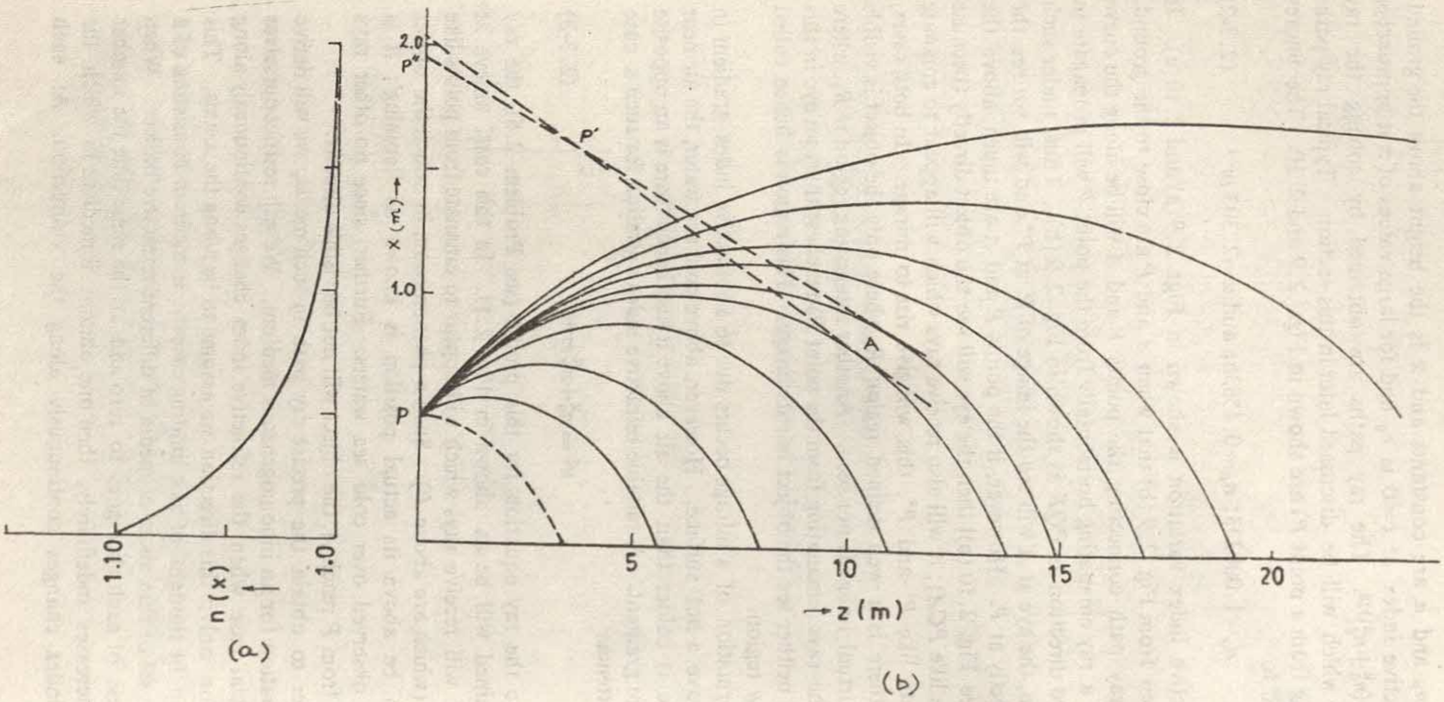


Fig. 2.11 Typical ray paths for an inhomogeneous medium characterized by Eq. (2.3–3) for an object point at a height of 0.5m. On the left is the refractive index variation.

interface, the light ray satisfies Snell's law and thus one would obtain

$$n_1 \sin \theta_1 = n_2 \sin \theta_2 = n_3 \sin \theta_3, \text{ etc.}$$

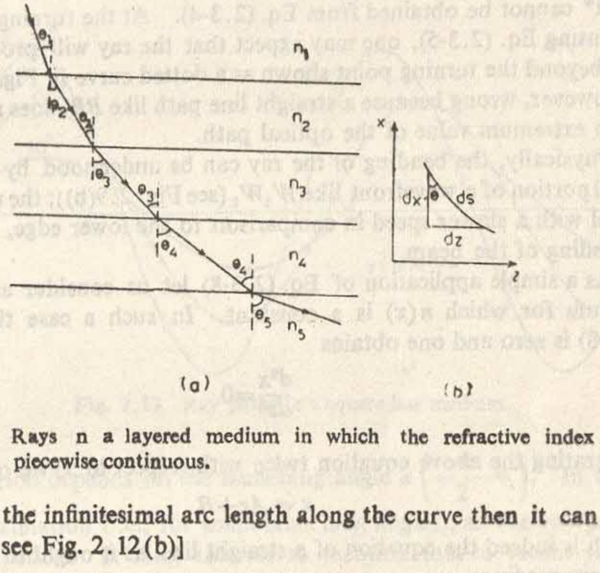
Thus we may put

$$n \sin \theta = \text{constant} \tag{2.3-4}$$

The constant can be determined from the fact that if the ray makes an angle θ_1 at a point where the refractive index is n_1, then the constant is $n_1 \sin \theta_1$. Thus

$$n(x) \sin \theta(x) = n_1 \sin \theta_1 \tag{2.3-5}$$

In the limiting case of continuous variation of the refractive index, the piecewise straight lines shown in Fig. 2.12 (a) form a continuous curve. If



(a) (b)

Fig. 2.12 Rays n a layered medium in which the refractive index is piecewise continuous.

ds represents the infinitesimal arc length along the curve then it can easily be seen that [see Fig. 2.12(b)]

$$(ds)^2 = (dx)^2 + (dz)^2 \tag{2.3-6}$$

If we refer to Fig. 2.12(b), we find

$$\frac{dz}{ds} = \sin \theta$$

Thus

$$\left(\frac{dx}{dz}\right)^2 + 1 = \left(\frac{ds}{dz}\right)^2 = \frac{1}{\sin^2 \theta}$$

or,

$$\left(\frac{dx}{dz}\right)^2 = \frac{n^2(x)}{n_1^2 \sin^2 \theta_1} - 1 \tag{2.3-7}$$

The solution of the above equation gives us the variation of x with z

for a given $n(x)$ which will give us the path of the ray in the medium with a refractive index variation $n(x)$. The above equation can be put in another form, by differentiating with respect to z,

$$2\frac{dx}{dz}\frac{d^2x}{dz^2}=\frac{1}{n_1^2\sin^2\theta_1}\frac{dn^2}{dx}\frac{dx}{dz}$$

or

$$\frac{d^2x}{dz^2}=\frac{1}{2\,n_1^2\sin^2\theta_1}\frac{dn^2(x)}{dx} \tag{2.3-8}$$

This equation is a special form of the more general equation which is known as the ray equation. Equations (2.3-7) and (2.3-8) and the ray equation can also be directly derived from Fermat's principle.

It should be pointed out that the bending up of the ray at the turning point* cannot be obtained from Eq. (2.3-4). At the turning point, $\theta=\pi/2$, and using Eq. (2.3-5), one may expect that the ray will proceed horizontally beyond the turning point shown as a dotted curve in Fig. 2.9(b). This is, however, wrong because a straight line path like BB' does not correspond to an extremum value of the optical path.

Physically, the bending of the ray can be understood by considering a small portion of a wavefront like W_1W_2 (see Fig. 2.9(b)); the upper edge will travel with a slower speed in comparison to the lower edge, this will cause a bending of the beam.

As a simple application of Eq. (2.3-8) let us consider a homogeneous medium for which $n(x)$ is a constant. In such a case the *RHS* of Eq. (2.3-8) is zero and one obtains

$$\frac{d^2x}{dz^2}=0$$

Integrating the above equation twice with respect to z, we obtain

$$x=Az+B \tag{2.3-9}$$

which is indeed the equation of a straight line, as it ought to be in a homogeneous medium.

Example 2.4 Solve the ray equation for following refractive index variation

$$n^2(x)=n_0^2-n_2x^2 \tag{2.3-10}$$

where n_0 represents the refractive index at $x=0$. (For $n_2x^2 \ll n_0^2$, $n(x)\simeq n_0-\frac{n_2}{2n_0}x^2$ which represents a parabolic variation of the refractive index. Such a medium is known as a square-law medium.)**

---

*The turning point corresponds to $dx/dz=0$; for example, the point B in Fig. 2.9(b).

**One has an approximate parabolic variation of refractive index in the D-layer of the ionosphere and one can use the solution to study the reflection of electromagnetic waves by the ionosphere (see Problem 8 in Ch. 6).

Solution. Substituting for n^2 from Eq. (2.3-10) in Eq. (2.3-8), we obtain

$$\frac{d^2x}{dz^2} = -\frac{n_2}{n_0^2 \sin^2\theta_1}\, x(z),$$

where θ_1 is the angle that the ray makes with x-axis at $x=0$, (n_0 is the refractive index at $x=0$). The solution of the above equation is

$$x(z) = A \sin\left[\frac{\sqrt{n_2}}{n_0 \sin\theta_1}\, z + \varphi_0\right] \tag{2.3-11}$$

where the constants A and φ_0 are to be determined from the initial position and slope of the ray (see also Problem 2.3). Typical ray paths are shown in Fig. 2.13. As can be seen from the figure, the ray paths are sinusoidal but the period depends on the launching angle $\alpha \left(=\frac{\pi}{2}-\theta_1\right)$. In the paraxial approximation (i.e., for small launching angles), all the ray paths have the same period. It is of interest to mention that in recent years, thin fibres with parabolic variation of refractive index have been fabricated and since all paraxial rays take approximately the same amount of time, they are characterized by extremely small pulse broadening. As such, fibres with square law variation of refractive index have tremendous potential use in optical communication [see Sodha and Ghatak (1977), Ghatak and Thyagarajan (1977)].

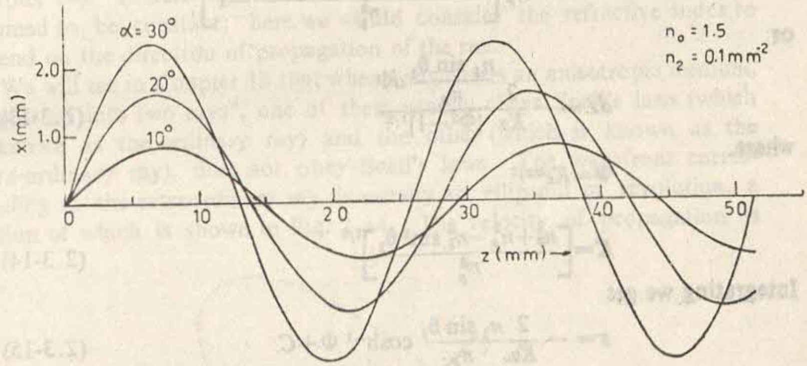


Fig. 2.13 Ray paths in a square law medium.

Example 2.5 Solve the ray equation [Eq. (2.3-7)], for the refractive index variation given by Eq. (2.3-1).

Solution. The refractive index distribution is

$$n^2 = n_0^2 + n_p^2[1 - e^{-\alpha x}] \tag{2.3-12}$$

Substituting this in Eq. (2.3-7), we obtain

$$\left(\frac{dx}{dz}\right)^2=\frac{n^2(x)}{n_1^2\sin^2\theta_1}-1$$

$$=\frac{n_0^2+n_p^2(1-e^{-\alpha x})}{n_1^2\sin^2\theta_1}-1$$

or

$$dz=\pm\frac{n_1\sin\theta_1\,dx}{[n_0^2+n_p^2(1-e^{-\alpha x})-n_1^2\sin^2\theta_1]^{1/2}},$$

where we have to take the negative sign for rays shown in Fig. 2.10 and a positive sign for rays travelling in the negative z direction. We rewrite the above equation in the form:

$$dz=-\frac{n_1\sin\theta_1\,e^{\alpha x/2}\,dx}{n_p\left[\dfrac{(n_0^2+n_p^2-n_1^2\sin^2\theta_1)\,e^{\alpha x}}{n_p^2}-1\right]^{1/2}}$$

or

$$dz=-\frac{2}{K\alpha}\frac{\dfrac{n_1\sin\theta_1}{n_p}\,d\Phi}{[\Phi^2-1]^{1/2}} \qquad (2.3\text{-}13)$$

where

$$\Phi=Ke^{\alpha x/2}$$

$$K=\left[\frac{n_0^2+n_p^2-n_1^2\sin^2\theta_1}{n_p^2}\right]^{1/2} \qquad (2.3\text{-}14)$$

Integrating we get

$$z=-\frac{2}{K\alpha}\frac{n_1\sin\theta_1}{n_p}\cosh^{-1}\Phi+C \qquad (2.3\text{-}15)$$

where C is a constant of integration and

$$\cosh\xi\equiv\tfrac{1}{2}(e^{\xi}+e^{-\xi}) \qquad (2.3\text{-}16)$$

Equation (2.3-15) can be rewritten in the form

$$\Phi=\cosh[\beta(C-z)] \qquad (2.3\text{-}17)$$

where

$$\beta=\frac{K\alpha\,n_p}{2n_1\sin\theta_1}.$$

Thus

$$Ke^{\alpha x/2}=\Phi=\cosh[\beta(c-z)]$$

or

$$x(z)=\frac{2}{\alpha}\ln\left[\frac{1}{K}\cosh\beta(C-z)\right] \qquad (2.3\text{-}18)$$

which gives the path of the ray. If at $z=0$, $x=x_0$, then

$$C=\frac{1}{\beta}\cosh^{-1}(Ke^{\alpha x_0/2}) \qquad (2.3\text{-}19,$$

For a given refractive index variation (i.e., for known values of n_0, n_p and α) and for a given position and slope of the ray (i.e. for known values of x_0 and θ_1) one can easily evaluate all the constants appearing in Eq. (2.3-18) and obtain the ray path. For the values of various parameters as given by Eq. (2.3-2) the explicit ray paths are plotted in Figs. 2.9 and 2.10.

2.4 Refraction of Rays at the Interface between an Isotropic and an Anisotropic Medium

We will now use Fermat's principle to study the refraction of a ray when it is incident from an isotropic homogeneous medium into an anisotropic homogeneous medium. In an anisotropic medium the velocity of propagation of light depends on the direction of propagation (see Chapter 18). In deriving Snell's laws in Sec. 2.2, the refractive index was assumed to be constant; here we would consider the refractive index to depend on the direction of propagation of the ray.

We will see in Chapter 18 that when a ray enters an anisotropic medium, it splits up into two rays*, one of them usually obeys Snell's laws (which is known as the ordinary ray) and the other (which is known as the extra-ordinary ray), does not obey Snell's laws. The wavefront corresponding to the extraordinary ray is usually an ellipsoid of revolution, a section of which is shown in Fig. 2.14. The velocity of propagation is

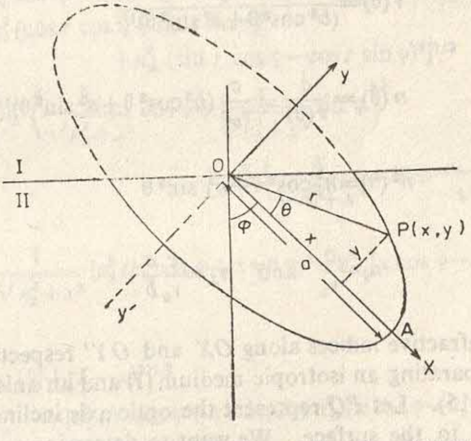


Fig. 2.14 The wavefronts corresponding to the extra-ordinary ray in an anisotropic medium may be an ellipsoid of revolution. A section of this ellipsoid (containing the optic axis) is shown in the above figure; *OX* represents the direction of the optic axis.

*It may be mentioned that the splitting of a ray into an ordinary and an extra-ordinary ray is not proved by Fermat's principle; here we are using Fermat's principle to find the path of the ray assuming a certain refractive index variation.

maximum along the direction OX and minimum along the direction OY'. In the particular case that we will be considering, the ellipsoid of revolution is obtained by rotating the ellipse about the OX axis. The speed of the ray in the general direction OP will be proportional to the length OP. Thus if $v(\theta)$ represents the speed of propagation in the general direction θ then*

$$v(\theta) = v_0 \frac{OP}{OA} = v_0 \frac{\sqrt{x^2+y^2}}{a} \tag{2.4-1}$$

where v_0 is the speed of propagation along $\theta=0$. Now, for an arbitrary point $P(x, y)$ on the ellipse,

$$\frac{x^2}{a^2}+\frac{y^2}{b^2}=1$$

If $OP=r$, then

$$x=r\cos\theta \quad \text{and} \quad y=r\sin\theta$$

Therefore,

$$r^2\left(\frac{\cos^2\theta}{a^2}+\frac{\sin^2\theta}{b^2}\right)=1$$

or

$$r=\frac{ab}{(b^2\cos^2\theta+a^2\sin^2\theta)^{1/2}} \tag{2.4-2}$$

Thus

$$v(\theta)=\frac{v_0 b}{(b^2\cos^2\theta+a^2\sin^2\theta)^{1/2}} \tag{2.4-3}$$

or

$$n(\theta)=\frac{c}{v(\theta)}=\frac{c}{v_0 b}(b^2\cos^2\theta+a^2\sin^2\theta)^{1/2} \tag{2.4-4}$$

Hence**

$$n^2(\theta)=n_{\parallel}^2\cos^2\theta+n_{\perp}^2\sin^2\theta \tag{2.4-5}$$

where

$$n_{\parallel}=\frac{c}{v_0} \quad \text{and} \quad n_{\perp}=\frac{c}{v_0}\frac{a}{b} \tag{2.4-6}$$

represent the refractive indices along OX and OY' respectively. Let S_1S_2 be a surface separating an isotropic medium (I) and an anisotropic medium (II) (see Fig. 2.15). Let PQ represent the optic axis inclined at an angle φ to the normal to the surface. We want to determine a relation between the angle of incidence i and the angle of refraction r. We will assume that the incident ray, refracted ray and the optic axis lie in the same plane.*

---

*The angle θ is being measured from OX, which is known as the optic axis. Along this direction the speed of the extra-ordinary ray is the same as that of the ordinary ray.

**It should be mentioned that $n(\theta)$ and $v(\theta)$ represent the ray refractive index and the ray velocity respectively; in Chapter 18, these are denoted by n_r and v_r respectively.

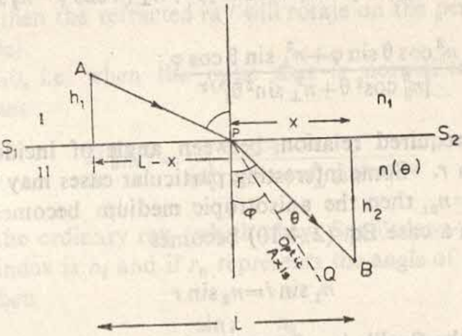


Fig. 2.15 Refraction of an incident ray when the optic axis lies in the plane of incidence.

The optical path length from *A* to *B* is

$$L_{op}=n_1\,[h_1^2+(L-x)^2]^{1/2}+n\,(\theta)\,[h_2^2+x^2]^{1/2} \tag{2.4-7}$$

where n_1 is the refractive index of medium I. Since $\theta=r-\varphi$, we have

$$\begin{aligned} n\,(\theta)&=[n_\parallel^2\cos^2(r-\varphi)+n_\perp^2\sin^2(r-\varphi)]^{1/2}\\ &=[n_\parallel^2(\cos r\cos\varphi+\sin r\sin\varphi)^2\\ &\qquad+n_\perp^2(\sin r\cos\varphi-\cos r\sin\varphi)^2]^{1/2}\\ &=\left[n_\parallel^2\left(\frac{h_2}{\sqrt{h_2^2+x^2}}\cos\varphi+\frac{x}{\sqrt{h_2^2+x^2}}\sin\varphi\right)^2\right.\\ &\qquad\left.+n_\perp^2\left(\frac{x}{\sqrt{h_2^2+x^2}}\cos\varphi-\frac{h_2}{\sqrt{h_2^2+x^2}}\sin\varphi\right)^2\right]^{1/2} \end{aligned}$$

or,

$$n\,(\theta)=\frac{1}{\sqrt{h_2^2+x^2}}\,[n_\parallel^2(h_2\cos\varphi+x\sin\varphi)^2+n_\perp^2(x\cos\varphi-h_2\sin\varphi)^2]^{1/2} \tag{2.4-8}$$

Thus

$$\begin{aligned} L_{op}=n_1\,[h_1^2+(L-x)^2]^{1/2}&\\ +[n_\parallel^2(h_2\cos\varphi+x\sin\varphi)^2&+n_\perp^2(x\cos\varphi-h_2\sin\varphi)^2]^{1/2} \end{aligned} \tag{2.4-9}$$

For the actual ray we must have

$$\frac{dL_{op}}{dx}=0$$

---

*In general, in an anisotropic medium the refracted ray does not lie in the plane of incidence However, it can be shown, from a three dimensional calculation (see Problem 2.9), that if the optic axis lies in the plane of incidence, then the refracted ray also lies in the plane of incidence. In the present calculation, we are assuming this and finding the direction of the refracted ray for a given angle of incidence.

i.e.,

$$\frac{n_1(L-x)}{[h_1^2+(L-x)^2]^{1/2}}=\frac{[n_\parallel^2(h_2\cos\varphi+x\sin\varphi)\sin\varphi+n_\perp^2(x\cos\varphi-h_2\sin\varphi)\cos\varphi]}{[n_\parallel^2(h_2\cos\varphi+x\sin\varphi)^2+n_\perp^2(x\cos\varphi-h_2\sin\varphi)^2]^{1/2}}$$

or,

$$n_1\sin i=\frac{n_\parallel^2\cos\theta\sin\varphi+n_\perp^2\sin\theta\cos\varphi}{[n_\parallel^2\cos^2\theta+n_\perp^2\sin^2\theta]^{1/2}} \tag{2.4-10}$$

which gives the required relation between angle of incidence i, and the angle of refraction r. Some interesting particular cases may be noted:

(i) when $n_\parallel=n_\perp=n_2$, then the anisotropic medium becomes an isotropic medium. In such a case Eq. (2.4-10) becomes

$$n_1\sin i=n_2\sin r \tag{2.4-11}$$

which is indeed the Snell's law for refraction at an interface between two isotropic media.

(ii) when $i=0$, Eq. (2.4-10) becomes

$$n_\parallel^2\cos\theta\sin\varphi+n_\perp^2\sin\theta\cos\varphi=0$$

or,

$$n_\parallel^2\cos(r-\varphi)\sin\varphi+n_\perp^2\sin(r-\varphi)\cos\varphi=0$$

or,

$$\cos r\,(n_\parallel^2-n_\perp^2)\cos\varphi\sin\varphi+\sin r\,(n_\parallel^2\sin^2\varphi+n_\perp^2\cos^2\varphi)=0$$

or,

$$\tan r=\frac{(n_\perp^2-n_\parallel^2)\cos\varphi\sin\varphi}{n_\parallel^2\sin^2\varphi+n_\perp^2\cos^2\varphi} \tag{2.4-12}$$

showing that, in general, $r\neq 0$. However, if $\varphi=\frac{\pi}{2}$ or 0 (i.e. if the optic axis is parallel to the surface or is normal to the surface), $r=0$ [see Figs. 2.16 (b) and (c)].

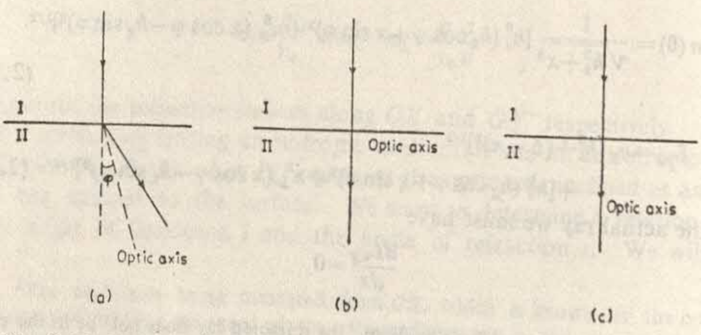


Fig. 2.16 In general, for a ray incident normally, the angle of refraction is not zero (a). However, if the optic axis is either parallel or perpendicular to the surface, then the incident ray goes undeviated as shown in (b) and (c).

Using Eq. (2.4-12) one can determine r for various values of φ. It can easily be seen that if the crystal is rotated about an axis which is normal to the surface then the refracted ray will rotate on the periphery of a cone (see Fig. 18.15c).

(iii) when $\varphi=0$, i.e. when the optic axis is normal to the surface, Eq. (2.4-10) becomes

$$\frac{n_1}{n_\parallel}\sin i=\frac{n_\parallel \cos r}{[n_\parallel^2\cos^2 r+n_\perp^2\sin^2 r]^{1/2}} \tag{2.4-13}$$

Now, for the ordinary ray (which obeys Snell's laws—see Chapter 18), the refractive index is $n_\parallel$ and if r_0 represents the angle of refraction for the ordinary ray then

$$\frac{\sin i}{\sin r_0}=\frac{n_\parallel}{n_1}$$

or

$$\sin r_0=\frac{n_1\sin i}{n_\parallel}=\frac{n_\parallel \cos r}{[n_\parallel^2\cos^2 r+n_\perp^2\sin^2 r]^{1/2}}$$

Thus

$$\tan r_0=\frac{\sin r_0}{(1-\sin^2 r_0)^{1/2}}=\frac{n_\parallel}{n_\perp}\tan r \tag{2.4-14}$$

All the above results can also be obtained by using Huygens' principle (see Chapter 18). It should be pointed out that in an anisotropic medium, the ray is *not*, in general, normal to the wavefront.

PROBLEMS

1. **If we rotate an ellipse about its major axis we will obtain, what is known as, an ellipsoid of revolution. Show by using Fermat's principle that all rays parallel to the major axis of the ellipse will focus to one of the focal points of the ellipse (see Fig. 2.17), provided the eccentricity of the ellipse equals n_1/n_2. (Hint: Start with the condition that $n_2AC'=n_1QB+n_2BC'$ and show that the point B (whose coordinates are x and y) lies on the periphery of an ellipse).**

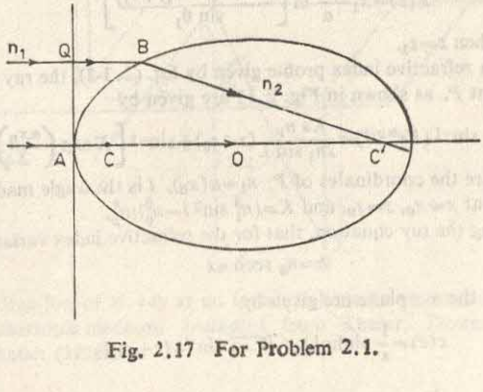


Fig. 2.17 For Problem 2.1.

2. C is the centre of a reflecting sphere of radius R (see Fig. 2.18). P_1 and P_2 are two points on a diameter equidistant from the centre. Obtain (a) the optical path length P_1O+P_2O as a function of θ and (b) find the values of θ for which P_1OP_2 is a ray path from reflection at the sphere.

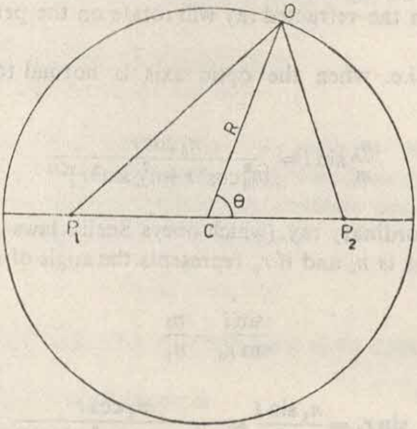


Fig. 2.18 A spherical reflector.

3. Since Eq. (2.3–8) has been derived by differentiating Eq. (2.3–7), only one of the constants in Eq. (2.3–11) is independent. Substitute Eq. (2.3–11) in Eq. (2.3–7) to show that

$$A=\frac{n_0 \cos \theta_1}{\sqrt{n_2}}=\frac{n_0}{\sqrt{n_2}} \sin \alpha$$

4. (a) In the above problem assuming $n_0=1.6$ and $n_2=0.1 \text{ mm}^{-2}$ calculate the paraxial focal length. (b) What will be the period of oscillation for paraxial rays and compare its value with the period corresponding to the launching angle $\alpha=4°$, 30° and 60°.
5. Consider a refractive index variation of the form

$$n=n_0\, e^{\alpha x}$$

Solve the ray equation and show that the ray path can be written in the form

$$x(z)=x_1-\frac{1}{\alpha} \ln\left[\frac{\sin\{\alpha(z-z_1)+\theta_1\}}{\sin \theta_1}\right]$$

where $x=x_1$ when $z=z_1$.
6. Show that for a refractive index profile given by Eq. (2.3-3), the ray paths emanating from a point P, as shown in Fig. 2.11 are given by

$$\sin^{-1}(Ke^{\alpha x/2})=\frac{K\alpha n_p}{2n_1 \sin i}(z-z_0)+\sin^{-1}\left[K\exp\left(\frac{\alpha x_0}{2}\right)\right]$$

where (z_0, x_0) are the coordinates of P, $n_1=n(x_0)$, i is the angle made by the ray with the x-axis at $x=x_0$, $z=z_0$, and $K=(n_1^2 \sin^2 i-n_0^2)/n_p^2$.
7. Show, by solving the ray equation, that for the refractive index variation

$$n=n_0 \operatorname{sech} \alpha x$$

the ray paths in the x-z plane are given by

$$x(z)=\frac{1}{\alpha} \sinh^{-1}[\sqrt{A^2-1} \sin\{\alpha(z+c)\}]$$

where A and C are constants to be determined from the initial conditions of the ray. Notice that the rays are periodic in z with a period which is independent of the initial conditions on the ray (cf. Example 2.4).

8. Consider a medium with refractive index variation of the form

$$n^2(x, y)=n_0^2[1-\alpha^2(x^2+y^2)]$$

where we are interested only in such values of x and y for which $\alpha^2(x^2+y^2)\ll 1$, (a medium characterized by the above refractive index variation finds widespread application in optical communications and integrated optics). For paraxial rays, i.e. rays which propagate close to the z-axis and which make small angles with this axis, the ray equation can be written in the form

$$\frac{d^2x}{dz^2}\simeq\frac{1}{2n_0^2}\frac{\partial n^2}{\partial x}$$

and

$$\frac{d^2y}{dz^2}\simeq\frac{1}{2n_0^2}\frac{\partial n^2}{\partial y}$$

where $x(z)$ and $y(z)$ represent the trajectory of the ray. Obtain a general solution of the above equations and show the existence of the helical ray (a helical ray spirals around the z-axis at a constant distance).

9. **Use Fermat's principle to show that on refraction at a plane interface between an isotropic medium of refractive index n_1 and a uniaxial medium of refractive index variation given by Eq. (2.4-5), the direction cosines of the incident ray (p_1, q_1, r_1), the refracted extra-ordinary ray (p, q, r) and the optic axis (l, m, n) are related by the following equations:**

$$\frac{n_\perp^2\, p+l(n_\perp^2-n_\parallel^2)\cos\theta}{(n_\perp^2\sin^2\theta+n_\parallel^2\cos^2\theta)^{1/2}}=-\,n_1p_1$$

and

$$\frac{n_\perp^2\, q+m(n_\perp^2-n_\parallel^2)\cos\theta}{(n_\perp^2\sin^2\theta+n_\parallel^2\cos^2\theta)^{1/2}}=-\,n_1q_1$$

where $\cos\theta=pl+qm+rn$ (choose the z-axis to be perpendicular to the plane interface—see Fig. 2.19. (Reference: Khular, Thyagarajan and Ghatak, 1976).

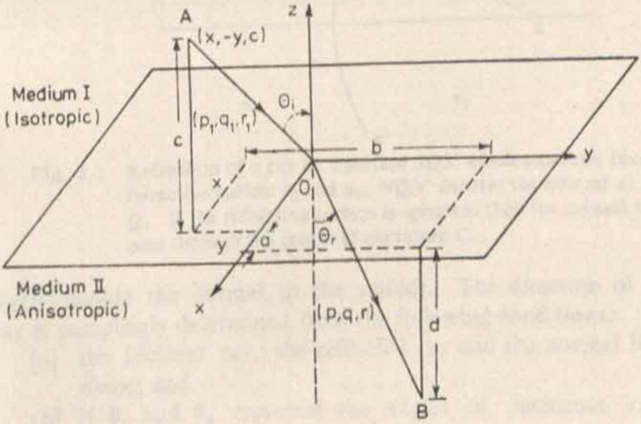


Fig. 2.19 Refraction of a ray at an interface between an isotropic and an anisotropic medium. [Adapted from Khular, Thyagarajan and Ghatak (1976)].

REFERENCES AND SUGGESTED READING

BORN, M. and WOLF, E. (1975), *Principles of Optics*, Pergamon Press, Oxford (England).

BUSH, R. T. and ROBINSON, R. S. (1974), "A note explaining the mirage," *Amer. J. Phys.*, 42, 774.

FEYNMANN, R. P., LEIGHTON, R. B. and SANDS, M. (1965), *The Feynman Lectures on Physics*, Vol. I, Addison-Wesley Publishing Co., Reading, Mass. (USA).

FRASER, A. B. and MACH, W. H. (1976), Mirages, *Scient. Amer.*, January, 234, 102.

GHATAK, A. K. and THYAGARAJAN, K. (1977), *Lasers and their Applications*, National Book Trust, India, New Delhi (in press).

HUMPHREYS, W. J. (1920), *Physics of the Air*, McGraw-Hill Book Co., New York (USA).

KHULAR, E., THYAGARAJAN, K. and GHATAK, A. K. (1977), A note on mirage formation, *Amer. J. Phys.*, 45, 90.

KHULAR, E., THYAGARAJAN, K. and GHATAK, A. K. (1976), Ray tracing in uniaxial and biaxial media, *Optik*, 46, 297.

PENFIELD, R. H. (1956), Consequences of parameter invariance in geometrical optics, *Amer. J. Phys.*, 24, 19.

SHAW, J. H. (1965), Fermat's principle and geometrical optics, *Amer. J. Phys.*, 33, 40.

SODHA, M. S., AGGARWAL, A. K. and KAW, P. K. (1967), Image formation by an optically stratified medium: Optics of mirage and looming, *Brit. J. Appl. Phys.*, 18, 503.

SODHA, M. S. and GHATAK A. K. (1977), *Inhomogeneous Optical Waveguides*, Plenum Publishing Corp., New York (USA).

3 Matrix Method in Paraxial Optics

> *The light dove, piercing in her easy flight through the air and perceiving its resistance, imagines that flight will be easier in empty space.*
>
> KANT

3.1. Introduction

Let us consider a ray *PQ* incident on a refracting surface *SQS'* separating two media of refractive indices n_1 and n_2 (see Fig. 3.1). Let *NQN'* denote the normal to the surface. The direction of the refracted ray is completely determined from the following conditions:

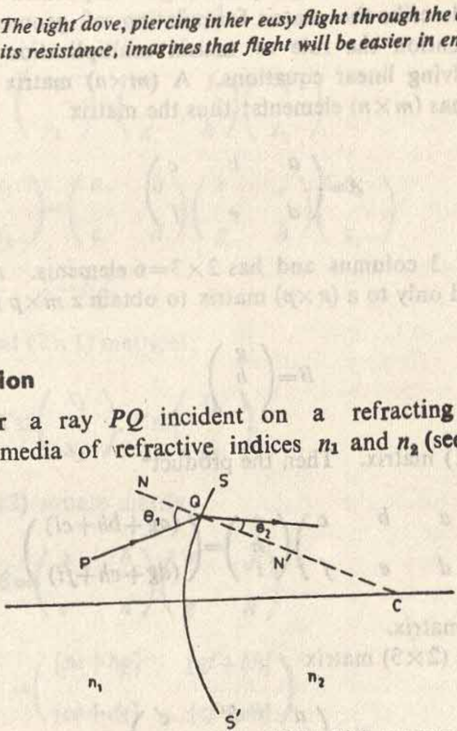


Fig. 3.1 Refraction of a ray by a surface *SQS'* which separates two media of refractive indices n_1 and n_2; *NQN'* denotes the normal at the point *Q*. If the refracting surface is spherical then the normal *NQN'* will pass through the centre of curvature *C*.

(a) the incident ray, the refracted ray and the normal lie in the same plane; and

(b) if θ_1 and θ_2 represent the angles of incidence and refraction respectively, then

$$\frac{\sin\theta_1}{\sin\theta_2}=\frac{n_2}{n_1} \qquad (3.1\text{-}1)$$

Optical systems, in general, are made up of a large number of refracting surfaces (like in a combination of lenses) and any ray can be traced through the system by using the above conditions. In order to obtain the position of the final image due to such a system, one has to calculate step by step the position of the image due to each surface and this image will act as an object for the next surface. Such a step by step analysis becomes complicated as the number of elements of an optical system increases. We shall, in this chapter, develop the matrix method which can be applied with ease under such situations. This method indeed lends itself to direct use in the computers for tracing rays through complicated optical systems.

Before we describe the matrix formulation of geometric optics it is necessary to mention the rule of matrix multiplication and the use of matrices for solving linear equations. A $(m\times n)$ matrix has m rows and n columns and has $(m\times n)$ elements; thus the matrix

$$A=\begin{pmatrix} a & b & c \\ d & e & f \end{pmatrix} \tag{3.1-2}$$

has 2 rows and 3 columns and has $2\times 3=6$ elements. A $(m\times n)$ matrix can be multiplied only to a $(n\times p)$ matrix to obtain a $m\times p$ matrix. Let

$$B=\begin{pmatrix} g \\ h \\ i \end{pmatrix} \tag{3.1-3}$$

represent a (3×1) matrix. Then the product*

$$AB=\begin{pmatrix} a & b & c \\ d & e & f \end{pmatrix}\begin{pmatrix} g \\ h \\ i \end{pmatrix}=\begin{pmatrix} (ag+bh+ci) \\ (dg+eh+fi) \end{pmatrix} \tag{3.1-4}$$

will be a (2×1) matrix.

If we define a (2×3) matrix

$$A'=\begin{pmatrix} a' & b' & c' \\ d' & e' & f' \end{pmatrix}$$

then

$$A'=A$$

if and only if $a'=a$, $b'=b$, $c'=c$, $d'=d$, $e'=e$ and $f'=f$, i.e. all the elements must be equal. Thus the set of two equations

$$\left.\begin{aligned} x_1&=ay_1+by_2 \\ x_2&=cy_1+dy_2 \end{aligned}\right\} \tag{3.1-5}$$

---

*In this case the product BA has no meaning.

can be written in the following form:

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}=\begin{pmatrix} (ay_1+by_2) \\ (cy_1+dy_2) \end{pmatrix}=\begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \tag{3.1-6}$$

the last step follows from the rule of matrix multiplication.

Further, if we have

and
$$\left.\begin{aligned} y_1&=ez_1+fz_2 \\ y_2&=gz_1+hz_2 \end{aligned}\right\} \tag{3.1-7}$$

then

$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix}=\begin{pmatrix} e & f \\ g & h \end{pmatrix}\begin{pmatrix} z_1 \\ z_2 \end{pmatrix} \tag{3.1-8}$$

Consequently,

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}=\begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} e & f \\ g & h \end{pmatrix}\begin{pmatrix} z_1 \\ z_2 \end{pmatrix} \tag{3.1-9}$$

or

$$X=BZ, \tag{3.1-10}$$

where X and Z represent (2×1) matrices:

$$X\equiv\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, Z\equiv\begin{pmatrix} z_1 \\ z_2 \end{pmatrix}, \tag{3.1-11}$$

and B represents a (2×2) square matrix

$$B=\begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} e & f \\ g & h \end{pmatrix}$$

$$=\begin{pmatrix} [ae+bg] & [af+bh] \\ [ce+dg] & [cf+dh] \end{pmatrix} \tag{3.1-12}$$

Equations (3.1-9) and (3.1-12) tell us that

and
$$\left.\begin{aligned} x_1&=(ae+bg)\,z_1+(af+bh)\,z_2 \\ x_2&=(ce+dg)\,z_1+(cf+dh)\,z_2 \end{aligned}\right\} \tag{3.1-13}$$

which can be verified by direct substitution.
We will now use the matrix method to trace paraxial rays through a cylindrically symmetric optical system.

3.2. The Matrix Method

We will consider a cylindrically symmetric optical system similar to the one shown in Fig. 3.2. The axis of symmetry is chosen as the z-axis. We

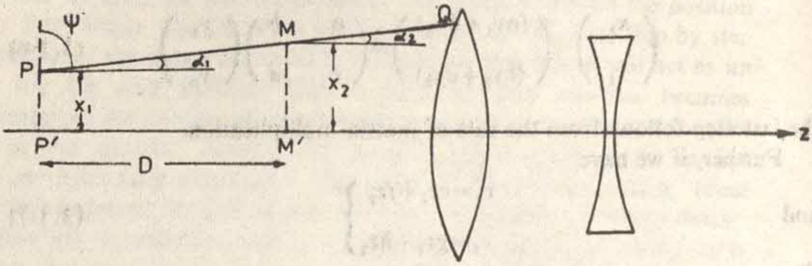


Fig. 3.2 In a homogeneous medium the ray travels in a straight line.

will be considering only paraxial rays in this chapter; non-paraxial rays lead to what are known as aberrations which will be discussed in the next chapter.

In the paraxial approximation we may confine ourselves to rays which pass through the axis of the system; these rays remain confined to a single plane. Such a ray can be specified by its distance from the axis of the system and the angle made by the ray with the axis; for example, in Fig. 3.2, the point P on the ray is at a distance x_1 from the axis and makes an angle α_1 with the axis. The quantities (x_1, α_1) represent the coordinates of the ray. However, instead of specifying the angle made by the ray with the z-axis, we will specify the quantity*

$$\lambda = n \cos \psi \ (= n \sin \alpha)$$

which represents the product of the refractive index with the cosine of the angle that the ray makes with the x-axis.

Now, when a ray propagates through an optical system, it undergoes only two operations; these are (a) translation and (b) refraction. The rays undergo translation when they propagate through a homogeneous medium as in the region PQ (see Fig. 3.2). However, when it strikes an interface of two media, it undergoes refraction. We will now study the effect of translation and of refraction on the coordinates of the ray.

(a) Effect of Translation

Consider a ray travelling in a homogeneous medium of refractive index n_1 which is initially at a distance x_1 from the z-axis and makes an angle α_1 with the axis (see point P in Fig. 3.2). Let (x_2, α_2) represent the coordinates of the ray at the point M (see Fig. 3.2). Since the medium is homogeneous, the ray travels in a straight line and, therefore,

$$\alpha_2 = \alpha_1. \tag{3.2-1}$$

Further, if PP' and MM' are perpendiculars on the axis and if $P'M' = D$, then

---

*This quantity is known as the optical direction cosine.

$$x_2 = x_1 + D \tan \alpha_1 \tag{3.2-2}$$

Since we are interested only in paraxial rays, α_1 is very small and hence we can make use of the approximation $\tan \alpha_1 \simeq \alpha_1$. Thus, Eq. (3.2-2) reduces to

$$x_2 \simeq x_1 + \alpha_1 D \tag{3.2-3}$$

If

$$\lambda_1 = n_1 \alpha_1 \tag{3.2-4}$$

and

$$\lambda_2 = n_2 \alpha_2 \tag{3.2-5}$$

then, using Eqs. (3.2-2) and (3.2-4) we get

$$\left.\begin{aligned} &\lambda_2 = \lambda_1, \\ \text{and} \qquad & \\ &x_2 = x_1 + \frac{D}{n_1} \lambda_1, \end{aligned}\right\} \tag{3.2-6}$$

which may be combined into the following matrix equation:

$$\begin{pmatrix} \lambda_2 \\ x_2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ D/n_1 & 1 \end{pmatrix} \begin{pmatrix} \lambda_1 \\ x_1 \end{pmatrix} \tag{3.2-7}$$

Thus, if a ray is initially specified by a (2×1) matrix with elements λ_1 and x_1, then the effect of translation through a distance D in a homogeneous medium of refractive index n_1, is completely given by the 2×2 matrix

$$T = \begin{pmatrix} 1 & 0 \\ D/n_1 & 1 \end{pmatrix} \tag{3.2-8}$$

and the final ray is given by Eq. (3.2-7). The matrix T is known as the translation matrix. Notice that

$$\det T = \begin{vmatrix} 1 & 0 \\ D/n_1 & 1 \end{vmatrix} = 1 \tag{3.2-9}$$

(b) Effect of Refraction

We will now determine the matrix which would represent the effect of refraction through a spherical surface of radius of curvature R. Consider the ray AP intersecting a spherical surface (separating two media of refractive indices n_1 and n_2 respectively) at the point P and getting refracted along PB (see Fig. 3.3). If θ_1 and θ_2 are the angles made by the incident and the refracted ray with the normal to the surface at P (i.e. with the line joining P to the centre of curvature C); then according to Snell's law

$$n_1 \sin \theta_1 = n_2 \sin \theta_2 \tag{3.2-10}$$

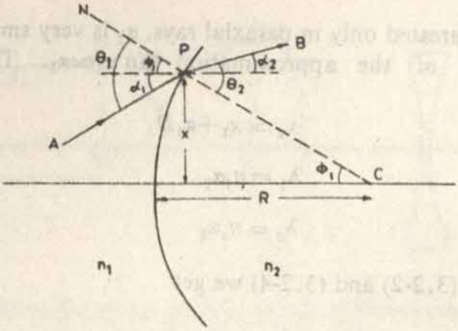


Fig. 3.3 The refraction of a ray at a spherical surface.

Since we are dealing with paraxial rays, one can make use of the approximation $\sin \theta \simeq \theta$. Thus Eq. (3.2-10) reduces to

$$n_1 \theta_1 \simeq n_2 \theta_2 \tag{3.2-11}$$

From Fig. 3.3 it follows that

$$\theta_1 = \varphi_1 + \alpha_1 \text{ and } \theta_2 = \varphi_1 + \alpha_2 \tag{3.2-12}$$

where α_1, α_2 and φ_1 are respectively the angles that the incident ray, the refracted ray and the normal to the surface make with the z-axis. Also, since φ_1 is small, we may write

$$\varphi_1 = \frac{x}{R} \tag{3.2-13}$$

Now, from Eqs. (3.2-11) and (3.2-12), we get

$$n_1 (\varphi_1 + \alpha_1) \simeq n_2 (\varphi_1 + \alpha_2)$$

or

$$n_2 \alpha_2 \simeq n_1 \alpha_1 - \frac{n_2 - n_1}{R} x \tag{3.2-14}$$

where we have used Eq. (3.3-13). Thus

$$\lambda_2 = \lambda_1 - Px \tag{3.2-15}$$

where

$$P = \frac{n_2 - n_1}{R} \tag{3.2-16}$$

is known as the power of the refracting surface. Also, since the height of the ray at P, before and after refraction is the same (i.e. $x_2 = x_1$) we obtain, for the refracted ray

$$\begin{pmatrix} \lambda_2 \\ x_2 \end{pmatrix} = \begin{pmatrix} 1 & -P \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \lambda_1 \\ x_1 \end{pmatrix} \tag{3.2-17}$$

Thus refraction through a spherical surface can be characterised by a 2×2 matrix:

$$\mathcal{R} = \begin{pmatrix} 1 & -P \\ 0 & 1 \end{pmatrix} \tag{3.2-18}$$

It may be noted here that

$$\det \mathcal{R} = \begin{vmatrix} 1 & -P \\ 0 & 1 \end{vmatrix} = 1 \tag{3.2-19}$$

In general, an optical system made up of a series of lenses, can be characterised by the refraction and translation matrices.

If a ray is specified by $\begin{pmatrix} \lambda_1 \\ x_1 \end{pmatrix}$ when it enters an optical system, and is specified by $\begin{pmatrix} \lambda_2 \\ x_2 \end{pmatrix}$ when it leaves the system, then one can, in general, write

$$\begin{pmatrix} \lambda_2 \\ x_2 \end{pmatrix} = \begin{pmatrix} b & -a \\ -d & c \end{pmatrix} \begin{pmatrix} \lambda_1 \\ x_1 \end{pmatrix} \tag{3.2-20}$$

where the matrix

$$S = \begin{pmatrix} b & -a \\ -d & c \end{pmatrix} \tag{3.2-21}$$

is called the system matrix and is determined solely by the optical system. The negative signs in some of the elements of S have been chosen for convenience. Since the only two operations a ray undergoes in traversing through an optical system are refraction and translation, the system matrix is, in general, a product of refraction and translation matrices. Also, using the property that the determinant of the product of matrices is the product of the determinant of the matrices, we obtain

$$\det S = 1 \tag{3.2-22}$$

i.e.

$$bc - ad = 1 \tag{3.2-23}$$

We will now derive the position of the image plane for an object plane, which is at a distance $(-D_1)$ from the first refracting surface of the optical system (see Fig. 3.4). Let the image be formed at a distance D_2 from the last refracting surface. We will be using the analytical geometry convention; i.e. for points on the left of a refracting surface, the distances will be negative and for points on the right of the refracting surface the distances will be positive, thus D_1 is an intrinsically negative quantity. Further, if D_2 is found to be positive the image is real and is

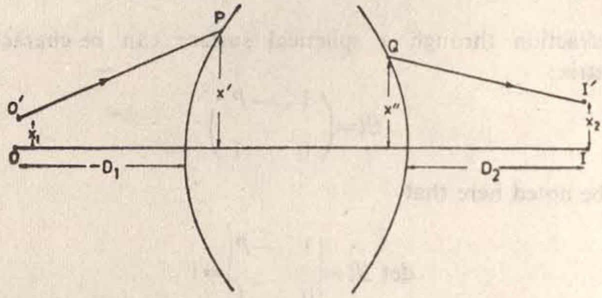


Fig. 3.4 The object point O is at a distance $(-D_1)$ from the first refracting surface. The paraxial image is assumed to be formed at a distance D_2 from the last refracting surface.

formed on the right of the refracting surface; on the other hand, if D_2 is found to be negative, the image will be virtual and will be formed on the left of the last refracting surface.

Let us consider a ray O'P starting from the point O' which lies in the object plane. Let QI' be the ray emerging from the last surface; the point I' is assumed to lie in the image plane—see Fig. 3.4 (the point I is the paraxial image of the point O and the image plane is defined to be the plane which contains the point I and is normal to the axis). Let (λ_1, x_1), (λ', x'), (λ'', x'') and (λ_2, x_2) represent the coordinates of the ray at O', P, Q and I' respectively. Then

$$\left.\begin{aligned} \begin{pmatrix} \lambda' \\ x' \end{pmatrix} &= \begin{pmatrix} 1 & 0 \\ -D_1 & 1 \end{pmatrix} \begin{pmatrix} \lambda_1 \\ x_1 \end{pmatrix} \\ \begin{pmatrix} \lambda'' \\ x'' \end{pmatrix} &= \begin{pmatrix} b & -a \\ -d & c \end{pmatrix} \begin{pmatrix} \lambda' \\ x' \end{pmatrix} \\ \begin{pmatrix} \lambda_2 \\ x_2 \end{pmatrix} &= \begin{pmatrix} 1 & 0 \\ D_2 & 1 \end{pmatrix} \begin{pmatrix} \lambda'' \\ x'' \end{pmatrix} \end{aligned}\right\} \tag{3.2-24}$$

Thus

$$\begin{pmatrix} \lambda_2 \\ x_2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ D_2 & 1 \end{pmatrix} \begin{pmatrix} b & -a \\ -d & c \end{pmatrix} \begin{pmatrix} 1 & 0 \\ -D_1 & 1 \end{pmatrix} \begin{pmatrix} \lambda_1 \\ x_1 \end{pmatrix} \tag{3.2-25}$$

where the first and the third matrices on the RHS correspond to translations by distances D_2 and $(-D_1)$ respectively (in a medium of refractive index unity); the second matrix correspond to the system matrix of the optical system. Carrying out the matrix multiplications, we obtain

$$\begin{pmatrix} \lambda_2 \\ x_2 \end{pmatrix} = \begin{pmatrix} b+aD_1 & -a \\ b\,D_2+a\,D_1D_2-cD_1-d & c-aD_2 \end{pmatrix} \begin{pmatrix} \lambda_1 \\ x_1 \end{pmatrix} \tag{3.2-26}$$

Thus

$$x_2=(bD_2+aD_1D_2-cD_1-d)\,\lambda_1+(c-aD_2)\,x_1 \tag{3.2-27}$$

For a ray emanating from the axial object point (i.e. for $x_1=0$) the image plane is determined by the condition $x_2=0$. Thus, for the image plane we must have

$$bD_2+aD_1D_2-cD_1-d=0 \tag{3.2-28}$$

which would give us the relationship between the distances D_1 and D_2. Thus, corresponding to the image plane, we have

$$\begin{pmatrix} \lambda_2 \\ x_2 \end{pmatrix} = \begin{pmatrix} b+aD_1 & -a \\ 0 & c-aD_2 \end{pmatrix} \begin{pmatrix} \lambda_1 \\ x_1 \end{pmatrix} \tag{3.2-29}$$

For $x_2\neq 0$, we obtain

$$x_2=(c-aD_2)\,x_1 \tag{3.2-30}$$

Consequently, the magnification of the system, $M\left(=\frac{x_2}{x_1}\right)$ would be given by

$$M=\frac{x_2}{x_1}=c-aD_2 \tag{3.2-31}$$

Further, since

$$\begin{vmatrix} b+ad_1 & -a \\ 0 & c-aD_2 \end{vmatrix}=1 \tag{3.2-32}$$

we obtain

$$b+aD_1=\frac{1}{c-aD_2}=\frac{1}{M} \tag{3.2-33}$$

Hence, if x_1 and x_2 correspond to object and image planes, then for a general optical system we may write

$$\begin{pmatrix} \lambda_2 \\ x_2 \end{pmatrix} = \begin{pmatrix} 1/M & -a \\ 0 & M \end{pmatrix} \begin{pmatrix} \lambda_1 \\ x_1 \end{pmatrix} \tag{3.2-34}$$

Example 3.1 **Obtain the system matrix for a thick lens and derive the thin lens and thick lens formulae.**

Solution. Let us consider a lens of thickness t and made of a material of relative refractive index n (see Fig. 3.5). Let R_1 and R_2 be the radii of curvatures of the two surfaces. The ray is assumed to strike the first surface of the lens at P and emerge from the point Q; let the coordinates of the ray at P and Q be

$$\begin{pmatrix} \lambda_1 \\ x_1 \end{pmatrix} \text{ and } \begin{pmatrix} \lambda_2 \\ x_2 \end{pmatrix} \tag{3.2-35}$$

where λ_1 and λ_2 are the optical direction cosines of the ray at P and Q, x_1 and x_2 are the distances of the points P and Q from the axis (see Fig. 3.5). The ray, in propagating from P to Q undergoes two refractions

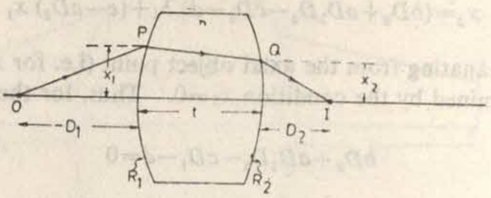


Fig. 3.5 A paraxial ray passing through a thick lens of thickness t.

[one at the first surface (whose radius of curvature is R_1) and the other at the second surface (whose radius of curvature is R_2)] and a translation through a distance* t in a medium of refractive index n. Thus

$$\begin{pmatrix} \lambda_2 \\ x_2 \end{pmatrix} = \begin{pmatrix} 1 & -P_2 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ t/n & 1 \end{pmatrix} \begin{pmatrix} 1 & -P_1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \lambda_1 \\ x_1 \end{pmatrix} \tag{3.2-36}$$

where

$$P_1 = \frac{n-1}{R_1} \quad \text{and} \quad P_2 = \frac{1-n}{R_2} = -\frac{n-1}{R_2} \tag{3.2-37}$$

represent the powers of the two refracting surfaces. Thus our system matrix is

$$S = \begin{pmatrix} b & -a \\ -d & c \end{pmatrix} = \begin{pmatrix} 1 & -P_2 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ t/n & 1 \end{pmatrix} \begin{pmatrix} 1 & -P_1 \\ 0 & 1 \end{pmatrix}$$

$$= \begin{pmatrix} 1 - \dfrac{P_2 t}{n} & -P_1 - P_2\left(1 - \dfrac{t}{n} P_1\right) \\ \dfrac{t}{n} & 1 - \dfrac{t}{n} P_1 \end{pmatrix} \tag{3.2-38}$$

For a thin lens, $t \to 0$ and the system matrix takes the following form

$$S = \begin{pmatrix} 1 & -P_1 - P_2 \\ 0 & 1 \end{pmatrix} \tag{3.2-39}$$

Thus for a thin lens,

$$a = P_1 + P_2, \; b = 1, \; c = 1 \quad \text{and} \quad d = 0 \tag{3.2-40}$$

Substituting the above values of a, b, c and d in Eq. (3.2-28), we obtain

---

*Notice that since we are dealing with paraxial rays, the distance between P and Q is approximately t.

$$D_2+(P_1+P_2)\,D_1D_2-D_1=0,$$

or

$$\frac{1}{D_2}-\frac{1}{D_1}=(P_1+P_2)=(n-1)\left(\frac{1}{R_1}-\frac{1}{R_2}\right) \tag{3.2-41}$$

or

$$\frac{1}{D_2}-\frac{1}{D_1}=\frac{1}{f} \tag{3.2-42}$$

where

$$f=\left[(n-1)\left(\frac{1}{R_1}-\frac{1}{R_2}\right)\right]^{-1} \tag{3.2-43}$$

represents the focal length of the lens. Equation (3.2-42) is the well-known thin lens formula. The signs of R_1 and R_2 for different kinds of lenses are shown in Fig. 3.6.

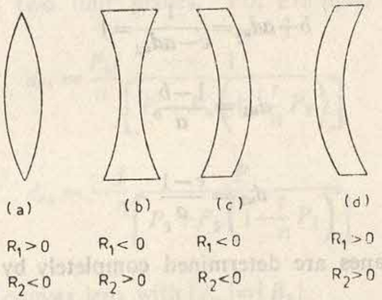


Fig. 3.6 Signs of R_1 and R_2 for different lens types.

For a thick lens, we have from Eq. (3.2-38)

$$\left.\begin{aligned} &a=P_1+P_2\left(1-\frac{t}{n}P_1\right),\; b=1-\frac{P_2t}{n} \\ &c=1-\frac{t}{n},\; d=\frac{t}{n} \end{aligned}\right\} \tag{3.2-44}$$

If we substitute the above values for a, b, c, and d in Eq. (3.2-28), we get the required relation between D_1 and D_2; however, for thick lenses it is more convenient to define the unit and the nodal planes which we shall do in the following sections.

3.3 Unit Planes

The unit planes are two planes, one each in the object and the image space, between which the magnification M is unity; i.e. any paraxial ray emanating from the unit plane in the object space will emerge at the same height from the unit plane in the image space. Thus, if d_{u1} and d_{u2} represent the distances of the unit planes from the refracting surfaces (see

Fig. 3.7)*, we obtain from Eq. (3.2-33):

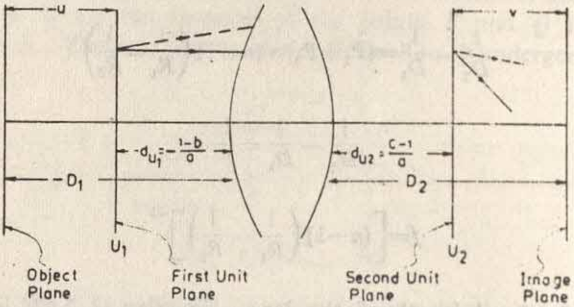


Fig. 3.7 U_1 and U_2 are the two unit planes. A ray emanating at any height from the first unit plane will cross the second unit plane at the same height.

$$b+ad_{u1}=\frac{1}{c-ad_{u2}}=1 \tag{3.3-1}$$

or

$$d_{u1}=\frac{1-b}{a}, \tag{3.3-2}$$

and

$$d_{u2}=\frac{c-1}{a} \tag{3.3-3}$$

Hence the unit planes are determined completely by the elements of the system matrix S.

It will be convenient to measure distances from the unit planes. Thus if u is the distance of the object plane from the first unit plane and v is the distance of the corresponding image plane from the second unit plane (see Fig. 3.7), we would obtain

$$D_1=u+d_{u1}=u+\frac{1-b}{a} \tag{3.3-4}$$

and

$$D_2=v+d_{u2}=v+\frac{c-1}{a} \tag{3.3-5}$$

Now, from Eq. (3.2-28) we have

$$D_2=\frac{d+cD_1}{b+aD_1} \tag{3.3-6}$$

Substituting for D_1 and D_2 from Eqs. (3.3-4) and (3.3-5) we get

$$v+\frac{c-1}{a}=\frac{d+cu+\frac{c(1-b)}{a}}{b+au+(1-b)}$$

---

*Obviously, if we consider U_1 as an object plane, then U_2 is the corresponding image plane.

or

$$v=\frac{ad-bc+c\,(au+1)-(c-1)\,(1+au)}{a\,(1+au)}$$

$$=\frac{au}{a\,(1+au)} \tag{3.3-7}$$

where we have used the condition that

$$\det S=bc-ad=1 \tag{3.3-8}$$

On simplification, we obtain

$$\frac{1}{v}-\frac{1}{u}=a \tag{3.3-9}$$

Thus $1/a$ represents the focal length of the system if the distances are measured from the two unit planes. For example, for a thick lens one obtains

$$d_{u1}=\frac{P_2 t}{n}\frac{1}{\left[P_1+P_2\left(1-\frac{t}{n}P_1\right)\right]} \tag{3.3-10}$$

and

$$d_{u2}=-\frac{t}{n}\frac{P}{\left[P_1+P_2\left(1-\frac{t}{n}P_1\right)\right]} \tag{3.3-11}$$

For a thick double convex lens with $|R_1|=|R_2|$

$$P_1=P_2=\frac{n-1}{R} \tag{3.3-12}$$

where $R=|R_1|=|R_2|$. Thus

$$d_{u1}=\frac{t}{n}\frac{1}{\left[2-\frac{t}{n}\frac{n-1}{R}\right]}\simeq\frac{t}{2n} \tag{3.3-13}$$

and

$$d_{u2}=-\frac{t}{n}\frac{1}{\left[2-\frac{t}{n}\frac{n-1}{R}\right]}\simeq-\frac{t}{2n} \tag{3.3-14}$$

where we have assumed $t \ll R$ which is indeed the case for most thick lenses. The positions of the unit planes are shown in Fig. 3.8. In order to calculate the focal length we note from Eq. (3.3-9) that

$$\frac{1}{f}=a=P_1+P_2\left(1-\frac{t}{n}P_1\right) \tag{3.3-15}$$

where we have used Eq. (3.2-44). Thus

$$\frac{1}{f}=(n-1)\left(\frac{1}{R_1}-\frac{1}{R_2}\right)+\frac{(n-1)^2 t}{nR_1R_2} \tag{3.3-16}$$

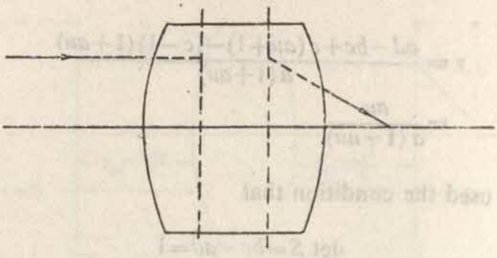


Fig. 3.8 Unit planes of a thick biconvex lens.

3.4 Nodal Planes

Nodal points are two points on the axis which have a relative angular magnification of unity, i.e. a ray striking the first point at an angle α emerges from the second point at the same angle (see Fig. 3.9). The planes which pass through these points and are normal to the axis are known as nodal planes.

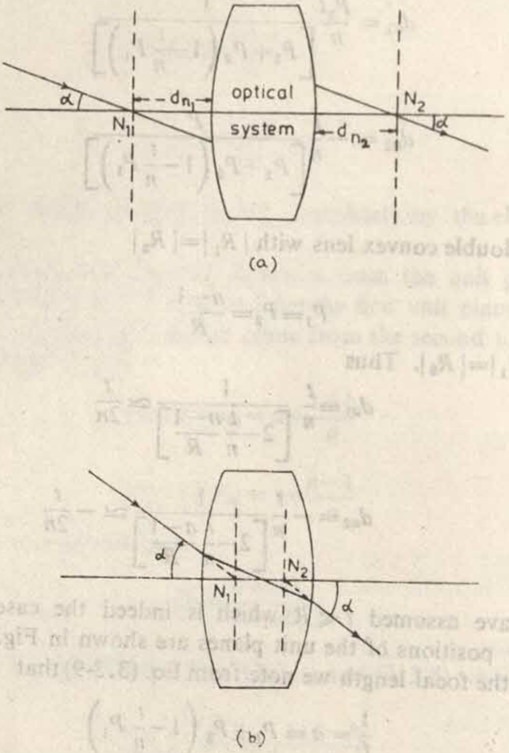


Fig. 3.9 N_1 and N_2 denote the two nodal points of an optical system. The nodal points can also lie inside the optical system as shown in (b).

To determine the position of the nodal points, we consider two axial points N_1 and N_2 at distances d_{n1} and d_{n2} from the two refracting surfaces respectively (see Fig. 3.9). From the definition of nodal points, we require that a ray incident at an angle α_1 on the point N_1 emerge from the optical system at the same angle α_1 from the other point N_2. Since we have assumed the media on either side of the system to have the same refractive index, this condition requires the equality of λ_1 and λ_2. Also, since we are considering an axial object point, $x_1=0$, we get from Eq. (3.2-29)

$$\lambda_2=(b+ad_{n1})\,\lambda_1=\lambda_1 \tag{3.4-1}$$

Thus

$$b+ad_{n1}=1 \tag{3.4-2}$$

or

$$d_{n1}=\frac{1-b}{a} \tag{3.4-3}$$

Comparing with Eq. (3.3-2) we find that $d_{n1}=d_{u1}$. This has arisen because of the equality of the indices of refraction on either side of the optical system. Similarly we can get

$$d_{n2}=\frac{c-1}{a} \tag{3.4-4}$$

Thus, when the media on either side of an optical system have the same refractive index (which is indeed the case for most optical systems), the nodal planes coincide with the unit planes. In general, if we know the elements of the system matrix S (i.e. if we know a, b, c and d, which are also called the Gaussian constants of the system), one can obtain all the properties of the system.

3.5. A System of Two Thin Lenses

We finally use the matrix formulation for the analysis of a combination of two thin lenses of focal lengths f_1 and f_2 separated by a distance t. The system matrix for the combination of the two lenses can be obtained by noting that the matrix of the two lenses are

$$\begin{pmatrix} 1 & -\frac{1}{f_1} \\ 0 & 1 \end{pmatrix} \text{ and } \begin{pmatrix} 1 & -\frac{1}{f_2} \\ 0 & 1 \end{pmatrix} \tag{3.5-1}$$

and the matrix for translation through a distance t (in air is) is

$$\begin{pmatrix} 1 & 0 \\ t & 1 \end{pmatrix} \tag{3.5-2}$$

Thus the system matrix S is

$$S=\begin{pmatrix} 1 & -\dfrac{1}{f_1} \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ t & 1 \end{pmatrix}\begin{pmatrix} 1 & -\dfrac{1}{f_2} \\ 0 & 1 \end{pmatrix}$$

$$=\begin{pmatrix} 1-\dfrac{t}{f_1} & -\dfrac{1}{f_1}-\dfrac{1}{f_2}+\dfrac{t}{f_1f_2} \\ t & 1-\dfrac{t}{f_2} \end{pmatrix} \tag{3.5-3}$$

As already noted, the element a in the system matrix represents the inverse of the focal length of the system. Thus the focal length of the combination would be

$$\frac{1}{f}=\frac{1}{f_1}+\frac{1}{f_2}-\frac{t}{f_1f_2}. \tag{3.5-4}$$

It is easy to see that if we have a system of four thin lenses we simply have to multiply seven matrices [four of them being of the type Eq. (3.5-1) and three of them of the type Eq. (3.5-2)].

Example 3.2 Consider a lens of thickness 1 cm, made of a material of refractive index 1.5, placed in air (refractive index of air $=1$). Let the radii of curvatures of the two surfaces be $+4$ cm and -4 cm (negative sign corresponds to a concave surface). Obtain the system matrix and determine the focal length and the positions of unit points and nodal points.

Solution. The power of the two surfaces are equal:

$$P_1=P_2=\frac{n-1}{R_1}=\frac{n-1}{|R_2|}=\frac{1.5-1}{4}=0.125\ \text{cm}^{-1}$$

Thus the system matrix is from Eq. (3.2-38),

$$\begin{pmatrix} 1-\dfrac{0.125\times 1}{1.5} & -0.125-0.125\left(1-\dfrac{1}{1.5}\times 0.125\right) \\ \dfrac{1}{1.5} & 1-\dfrac{0.125}{1.5} \end{pmatrix}$$

$$=\begin{pmatrix} 0.9167 & -0.240 \\ 0.6667 & 0.9167 \end{pmatrix}$$

Thus the focal length of the lens would be

$$f=\frac{1}{0.24}\simeq 4.2\ \text{cm}$$

The position of the unit planes are given by [see Eqs. (3.3-2) and (3.3-3)]

$$d_{u1}=\frac{1-b}{a}\quad\frac{1-0.9167}{0.24}=\frac{0.0833}{0.24}\approx 0.35\text{ cm}$$

$$d_{u2}=\frac{c-1}{a}=\frac{0.9167-1}{0.24}=-\frac{0.0833}{0.24}\approx -0.35\text{ cm}$$

Notice that d_{u1} is positive and d_{u2} is negative. Thus the unit planes are situated as shown in Fig. 3.8. The nodal points coincide with the unit points because the lens is immersed in a homogeneous medium.

Example 3.3 Consider a system of two thin lenses as shown in Fig. 3.10. For a 1 cm tall object at a distance of 40 cm from the convex lens, calculate the position and size of the image.

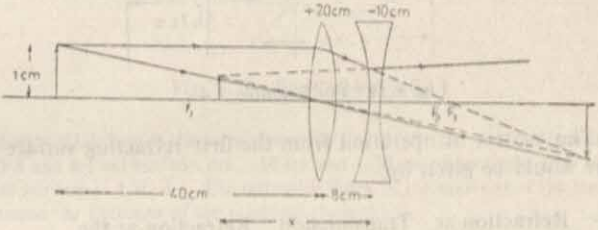


Fig. 3.10 For Example 3.3.

Solution. Let v be the distance of the image plane from the concave lens. Thus the matrix, which when operated on the object column matrix gives the image column matrix, would be given by

$$\begin{array}{ccccc}
\text{Concave lens to image} & \text{Concave lens} & \text{Convex lens to concav lens} & \text{Convex lule} & \text{Object to convex lens} \\
\begin{pmatrix} 1 & 0 \\ v & 1 \end{pmatrix} & \begin{pmatrix} 1 & +\frac{1}{10} \\ 0 & 1 \end{pmatrix} & \begin{pmatrix} 1 & 0 \\ 8 & 1 \end{pmatrix} & \begin{pmatrix} 1 & -\frac{1}{20} \\ 0 & 1 \end{pmatrix} & \begin{pmatrix} 1 & 0 \\ 40 & 1 \end{pmatrix}
\end{array}$$

$$=\begin{pmatrix} 1 & 0 \\ v & 1 \end{pmatrix}\begin{pmatrix} 2.2 & 0.01 \\ +32 & 0.6 \end{pmatrix}=\begin{pmatrix} 2.2 & 0.01 \\ 2.2v+32 & 0.6+0.01v \end{pmatrix}$$

The image plane would correspond to

$$32+2.2v=0$$

or

$$v\simeq -14.5\text{ cm}$$

i.e. it is at a distance of 14.5 cm on the left of the concave lens. If we compare with Eq. (3.2-24), we obtain

$$M=0.6+0.01v=0.6-0.01\frac{32}{2.2}=+\frac{1}{2.2}$$

Thus the magnification is 1/2.2 (which is positive producing a virtual image).

Example 3.4 Consider a sphere of radius 20 cm of refractive index 1.6 (see Fig. 3.11). Find the position of the paraxial focal point.

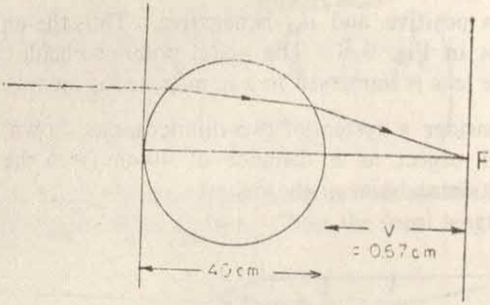


Fig. 3.11 For Example 3.4.

Solution. The matrix of operation from the first refracting surface to the image plane would be given by:

Second surface to Image	Refraction at second surface	Transmission through glass	Refraction at the first surface

$$\begin{pmatrix}1 & 0\\ v & 1\end{pmatrix}\begin{pmatrix}1 & +\frac{1-1.6}{20}\\ 0 & 1\end{pmatrix}\begin{pmatrix}1 & 0\\ \frac{40}{1.6} & 1\end{pmatrix}\begin{pmatrix}1 & -\frac{(1.6-1)}{20}\\ 0 & 1\end{pmatrix}$$

$$=\begin{pmatrix}1 & 0\\ v & 1\end{pmatrix}\begin{pmatrix}0.25 & -0.0375\\ 25 & 0.25\end{pmatrix}$$

$$=\begin{pmatrix}0.25 & -0.0375\\ 25+0.25v & 0.25-0.0375v\end{pmatrix}$$

Thus at the image plane the ray coordinates would be

$$\begin{pmatrix}\lambda_2\\ x_2\end{pmatrix}=\begin{pmatrix}0.25 & -0.0375\\ 25+0.25v & 0.25-0.0375v\end{pmatrix}\begin{pmatrix}\lambda_1\\ x_1\end{pmatrix}$$

This gives us

$$x_2=(25+0.25v)\lambda_1+(0.25-0.0375v)x_1$$

To determine the focal distance v consider a ray incident parallel to the axis for which $\lambda_1=0$. The focal plane would be that plane for which x_2 is also zero. This gives us

$$0.375v=0.25 \quad \text{or} \quad v=0.67\,\text{cm}$$

PROBLEMS

1. Starting with any three 2×2 matrices A, B and C show explicitly that $(AB)\,C = A\,(BC)$, i.e. matrices satisfy the associative law of multiplication.
2. Consider a hemisphere of radius 20 cm and refractive index 1.5. If H_1 and H_2 denote the positions of the first and second principal points then show that $AH_1 = 13.3$ cm and that H_2 lies on the second surface as shown in Fig. 3.12. Further, the focal length is 40 cm.

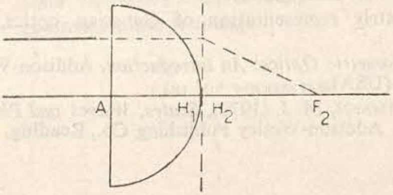


Fig. 3.12 For Problem 3.2.

3 Consider a thick lens of the form shown in Fig. 3.13; the radii of curvatures of the first and second surfaces are -10 cm and $+20$ cm respectively and the thickness of the lens is 1.0 cm. The refractive index of the material of the lens is 1.5. Determine the positions of the principal planes.

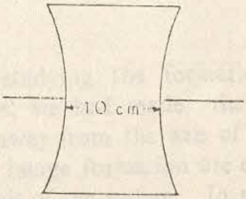


Fig. 3.13 Sketch of a thick double concave lens.

4. Consider a combination of two thin lenses of focal lengths f_1 and f_2 separated by a distance (f_1+f_2). Show that the angular magnification of the lens combination $\left(\text{which is just } \frac{\lambda_2}{\lambda_1} = \frac{\alpha_2}{\alpha_1}\right)$ is given by $-f_1/f_2$. Interpret the negative sign in the expression for magnification.
5. Consider a thin convex lens of focal length f. Suppose that an object of height h is placed at a distance slightly less than f from the lens. Will the image be real or virtual? Show that the angle subtended by image at the lens would be h/f. (In order to view the object comfortably, the object would have to be at a distance of 25 cm and the angle subtended by the object would be $h/25$. Thus the ratio $(h/f)/(h/25) = 25/f$ represents the magnification).
6. Consider a plane glass plate of thickness 2 cm made of a material of refractive index 1.5. Find the position of the image of an object placed at a distance of 5 cm from it.
7. Consider an object lying at the bottom of a swimming pool of water of height 6 m. Assuming the refractive index of water to be 4/3, find the apparent position of the object when viewed from above the pool.

REFERENCES AND SUGGESTED READING

BLAKER, J. N. (1971), *Geometrical Optics: The Matrix Theory,* Marcel Dekker, New York (USA).

BROUWER, W. (1964), *Matrix Methods in Optical Instrumental Design,* Benjamin, New York (USA).

EAKIN, D. M. and DAVIS, S. P. (1966), An application of matrix optics, *Amer. J. Phys.*, 34, 758.

GERRARD, A. and BURCH, J. M. (1975), *Introduction to Matrix Methods in Optics,* John Wiley & Sons, New York (USA).

HALBACH, K. (1964), Matrix representation of Gaussian optics, *Amer. J. Phys.*, 32, 90.

NUSSBAUM, A. (1968), *Geometric Optics: An Introduction,* Addison-Wesley Publishing Co., Reading, Mass. (USA).

SIMMONS, J. W. and GUTTMANN, M. J. (1970), *States, Waves and Photons: A Modern Introduction to Light,* Addison-Wesley Publishing Co., Reading, Mass. (USA).

4 Aberrations

> *Since 'tis Nature's law to change*
> *constancy alone is strange*
>
> EARL OF ROCHESTER

4.1 Introduction

In Chapter 3, while studying the formation of images by refracting surfaces and thin lenses, we had made the assumption that the object point does not lie far away from the axis of the optical system and that the rays taking part in image formation are essentially those which make small angles with the axis of the system. In practice, neither of the above assumptions is true; one in fact has to deal with rays making large angles with the axis. The domain of optics dealing with rays lying close to the optical axis and making small angles with it is called paraxial optics. We had found that in the realm of paraxial optics, the images of objects were perfect, i.e., all rays emanating from a single object point converged to a single image point and the magnification of the system was a constant of the optical system, independent of the particular ray under consideration. Since in real optical systems, nonparaxial rays also take part in image formation, the actual images depart from the ideal images. This departure leads to what are known as aberrations.

It can be shown that the primary aberrations of any rotationally symmetric system can be specified by five coefficients. The five coefficients represent the spherical aberration, coma, astigmatism, curvature of field and distortion. These are called the Seidel aberrations. Since these aberrations are present even for light of a single wavelength, they are also called monochromatic aberrations. In this chapter, we will consider the five kinds of aberrations separately and discuss the effect on the image when each one of them is present separately.

It should be mentioned that if a polychromatic source (like white light) is used for image formation (which is indeed the case for many optical instruments) then, in general, the image will be coloured; this is known as chromatic aberration. Physically, chromatic aberration is due to the dependence of the refractive index of the material of the lens on wavelength of the radiation under consideration. Since image formation is accompanied by refraction at refractive index discontinuities, the wavelength dependence of the refractive index results in the coloured image. For a polychromatic source, different wavelength components (after refraction) proceed along different directions and form images at different points; this leads to coloured images. Since chromatic aberration is the easiest to understand, we would discuss this first. This will be followed by a discussion of monochromatic aberrations.

4.2 Chromatic Aberration

Let us consider a parallel beam of white light incident on a thin convex lens as shown in Fig. 4.1. Since blue light gets refracted more than red light, the point at which the blue light would focus is nearer the lens than the point at which the red light would focus. Thus, the image will appear to be coloured; it may be mentioned that this aberration is independent of the five Seidel aberrations to be discussed in later sections.

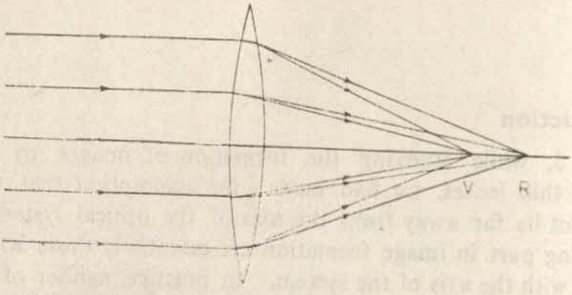


Fig. 4.1 When white light consisting of a continuous range of wavelengths is incident on a lens, then each wavelength refracts by different amounts; this leads to chromatic aberration. This aberration is independent of the five Seidel aberrations.

For the case of a thin lens the expression for chromatic aberration can easily be derived. The focal length of a thin lens is given by

$$\frac{1}{f}=(n-1)\left(\frac{1}{R_1}-\frac{1}{R_2}\right) \tag{4.2-1}$$

If a change of n by δn (the change of n is due to the change in the wavelength of the light) results in a change of f by δf then we obtain by differentiating Eq. (4.2-1),

$$-\frac{\delta f}{f^2} = \delta n\left(\frac{1}{R_1}-\frac{1}{R_2}\right) = \frac{\delta n}{n-1}\,\frac{1}{f}$$

i.e.

$$\delta f = -f\,\frac{\delta n}{n-1} \tag{4.2-2}$$

which represents the chromatic aberration of a thin lens. If n_b and n_r represent the refractive indices for the blue and red colours respectively then

$$f_r - f_b = f\left(\frac{n_b - n_r}{n-1}\right) \tag{4.2-3}$$

would represent the chromatic aberration.

4.2.1 The Achromatic Doublet

We will first consider an optical system of two thin lenses made of different materials placed in contact with each other. For example, one of the lenses may be made of crown glass and the other of flint glass. We will find the condition for this lens combination to have the same focal length for the blue and red colours. Let n_b, n_y and n_r represent the refractive indices for the material of the first lens corresponding to the blue, yellow and red colours respectively. Similarly, n'_b, n'_y and n'_r represent the corresponding refractive indices for the second lens. If f_b and f'_b represent the focal lengths for the first and the second lens corresponding to the blue colour, and if F_b represents the focal length of the combination of the two lenses (placed in contact), then

$$\begin{aligned}\frac{1}{F_b} &= \frac{1}{f_b}+\frac{1}{f'_b}\\ &= (n_b-1)\left(\frac{1}{R_1}-\frac{1}{R_2}\right)+(n'_b-1)\left(\frac{1}{R'_1}-\frac{1}{R'_2}\right)\end{aligned} \tag{4.2-4}$$

where R_1 and R_2 represent the radii of curvatures of the first and second surface for the first lens and, as before, the primed quantities refer to the second lens. Thus, we may write

$$\frac{1}{F_b} = \frac{n_b-1}{n-1}\,\frac{1}{f}+\frac{n'_b-1}{n'-1}\,\frac{1}{f'} \tag{4.2-5}$$

where

$$\frac{1}{f} \equiv (n-1)\left(\frac{1}{R_1}-\frac{1}{R_2}\right),\ \frac{1}{f'} \equiv (n'-1)\left(\frac{1}{R'_1}-\frac{1}{R'_2}\right) \tag{4.2-6}$$

$$n \equiv \frac{n_b+n_r}{2} \approx n_y,\ n' \equiv \frac{n'_b+n'_r}{2} \approx n'_y \tag{4.2-7}$$

f and f' represent the focal lengths of the first and second lens corresponding to a mean colour which is around the yellow region. Similarly, the focal length of the combination corresponding to the red colour would be given by

$$\frac{1}{F_r}=\frac{n_r-1}{n-1}\frac{1}{f}+\frac{n'_r-1}{n'-1}\frac{1}{f'} \tag{4.2-8}$$

For the focal length of the combination to be equal for blue and red colours, we must have

$$\frac{n_b-1}{n-1}\frac{1}{f}+\frac{n'_b-1}{n'-1}\frac{1}{f'}=\frac{n_r-1}{n-1}\frac{1}{f}+\frac{n'_r-1}{n'-1}\frac{1}{f'}$$

or

$$\frac{\omega}{f}+\frac{\omega'}{f'}=0 \tag{4.2-9}$$

where

$$\omega=\frac{n_b-n_r}{n-1} \text{ and } \omega'=\frac{n'_b-n'_r}{n'-1} \tag{4.2-10}$$

are known as the dispersive powers. Since ω and ω' are both positive, f and f' must be of opposite signs for the validity of Eq. (4.2-9). A lens combination which satisfies Eq. (4.2-9) is known as an achromatic doublet (see Fig. 4.2). It may be mentioned that if the two lenses are made of the same material, then $\omega=\omega'$ and Eq. (4.2-9) would imply $f=-f'$; such a combination will have an infinite focal length. Thus, for an achromatic doublet the two lenses must be of different materials.

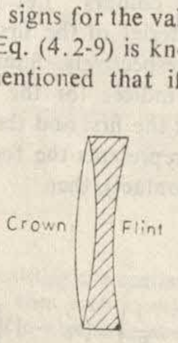


Fig. 4.2 An achromatic doublet.

Example 4.1 An achromatic doublet of focal length 20 cm is to be made by placing a convex lens of borosilicate crown glass in contact with a diverging lens of dense flint glass. Assuming $n_r=1.51462$, $n_b=1.52264$, $n'_r=1.61216$ and $n'_b=1.62901$, calculate the focal length of each lens; here the unprimed and the primed quantities refer to the borosilicate crown glass and dense flint glass respectively.

Solution.

$$n\simeq\frac{n_b+n_r}{2}=\frac{1.52264+1.51462}{2}$$
$$=1.51863$$

$$n' \simeq \frac{n_b'+n_r'}{2} = \frac{1.62901+1.61216}{2}$$

$$=1.62058$$

Thus,

$$\omega = \frac{1.52264-1.51462}{1.51863-1} = 0.01546$$

and

$$\omega' = \frac{1.62901-1.61216}{1.62058-1} = 0.02715$$

substituting in Eq. (4.2-9), we obtain

$$\frac{0.01546}{f} + \frac{0.02715}{f'} = 0$$

or

$$\frac{f}{f'} = -0.56942$$

Now, for the lens combination to be of focal length 20 cm we must have

$$\frac{1}{f} + \frac{1}{f'} = \frac{1}{20}$$

or

$$\frac{1}{f}\Big[1-0.56942\Big] = \frac{1}{20}$$

or

$$f = 20 \times 0.43058 = 8.61 \text{ cm.}$$

and

$$f' = -\frac{f}{0.56942} \simeq -15.1 \text{ cm.}$$

4.2.2 Removal of Chromatic Aberration of a Separated Doublet

Let us consider two thin lenses of focal lengths f and f' and separated by a distance t (see Fig. 4.3). The focal length of the combination F,

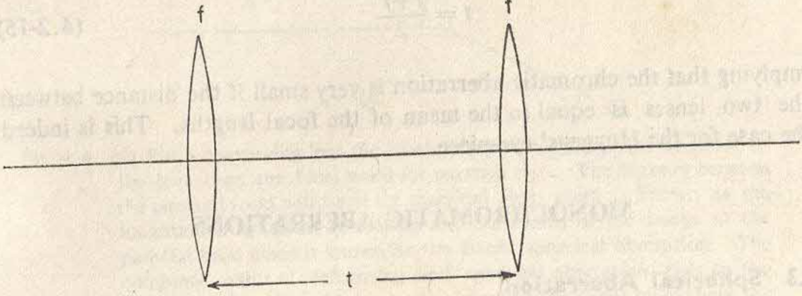


Fig. 4.3 The separated doublet.

would be

$$\frac{1}{F}=\frac{1}{f}+\frac{1}{f'}-\frac{t}{ff'} \tag{4.2-11}$$

The focal length of the first lens would be given by

$$\frac{1}{f}=(n-1)\left(\frac{1}{R_1}-\frac{1}{R_2}\right) \tag{4.2-12}$$

with a similar expression for $1/f'$. If Δf and Δn represent the changes in the focal length and in the refractive index due to a change $\Delta\lambda$ in the wavelength, then by differentiating Eq. (4.2-12) we obtain

$$-\frac{\Delta f}{f^2}=\Delta n\left(\frac{1}{R_1}-\frac{1}{R_2}\right)=\frac{\Delta n}{(n-1)f}$$

Thus, differentiating Eq. (4.2-11), we obtain

$$\begin{aligned}-\frac{\Delta F}{F^2}&=-\frac{\Delta f}{f^2}-\frac{\Delta f'}{f'^2}+\frac{t}{f}\frac{\Delta f'}{f'^2}+\frac{t}{f'}\frac{\Delta f}{f^2}\\&=\frac{\Delta n}{(n-1)f}+\frac{\Delta n'}{(n'-1)f'}-\frac{t}{f}\frac{\Delta n'}{(n'-1)f'}-\frac{t}{f'}\frac{\Delta n}{(n-1)f}\\&=\frac{\omega}{f}+\frac{\omega'}{f'}-\frac{t}{ff'}(\omega+\omega')\end{aligned} \tag{4.2-13}$$

where, as before, ω and ω' represent the dispersive powers. Consequently, for the combination to have the same focal length for blue and red colours we should have

$$\frac{t(\omega+\omega')}{ff'}=\frac{\omega}{f}+\frac{\omega'}{f'}$$

or

$$t=\frac{\omega f'+\omega' f}{\omega+\omega'} \tag{4.2-14}$$

If both the lenses are made of the same material, then $\omega=\omega'$ and the above equation simplifies to

$$t=\frac{f+f'}{2} \tag{4.2-15}$$

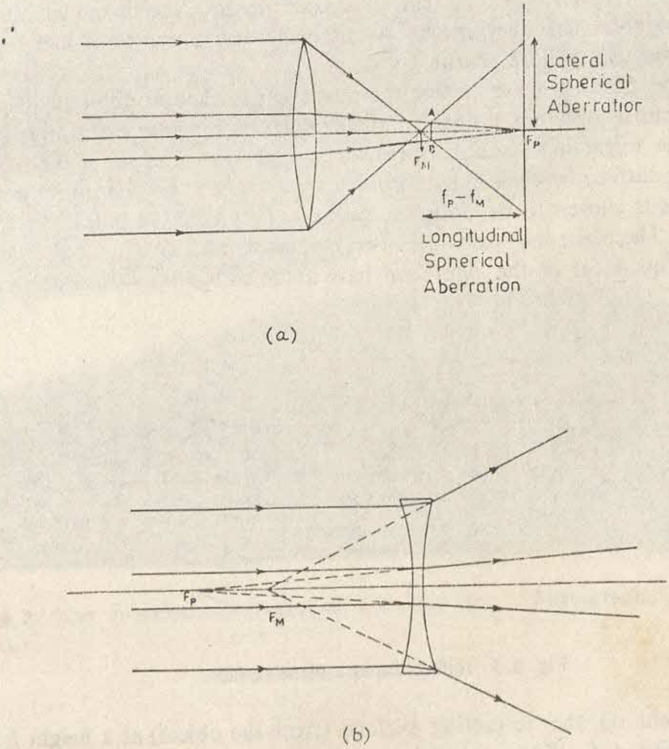
implying that the chromatic aberration is very small if the distance between the two lenses is equal to the mean of the focal lengths. This is indeed the case for the Huygens' eye-piece.

MONOCHROMATIC ABERRATIONS

4.3 Spherical Aberration

Let a beam of light parallel to the axis be incident on a thin lens (see

Fig. 4.4). The light rays after passing the lens bend towards the axis and cross the axis at some point. If we restrict ourselves to the paraxial region, then we can see that all rays cross the z-axis at the same point which is at a distance f_p from the lens; f_p represents the paraxial focal length of the lens. If one does not restrict to the paraxial region, then in general, rays which are incident at different heights on the lens, hit the axis at different points. For example, for a convex lens, the marginal rays (which are incident near the periphery of the lens) focus at a point closer than the focal point of paraxial rays (see Fig. 4.4a). Similarly, for a concave lens, rays which are incident farther from the axis appear to be emerging from a point which is nearer to the lens (see Fig. 4.4b). The point at which the paraxial rays strike the axis (F_P) is called the paraxial focus and the point at which the rays near the periphery strike is called the marginal focus



Fg. 4.4 (a) For a converging lens the focal point for marginal rays lie closer to the lens than the focal point for paraxial rays. The distance between the paraxial focal point and the marginal focal point is known as the longitudinal spherical aberration and the radius of the image at the paraxial focal plane is known as the lateral spherical aberration. The combined effect of defocusing and spherical aberration leads to the formation of a circle of least confusion, where the image would have the minimum diameter. (b) The spherical aberration of a diverging lens.

(F_M). The distance between the two foci is a measure of spherical aberration in the lens. Thus if O represents an axial object, then different rays emerging from the object converge to different points; consequently, the image of a point object will not be a point. The distance along the axis between the paraxial image point and the image corresponding to marginal rays (i.e. rays striking the edge of the lens) is termed longitudinal spherical aberration. Similarly, the distance between the paraxial image point and the point at which the marginal ray strikes the paraxial image plane is called the lateral spherical aberration (see Fig. 4.4a). The image on any plane (normal to the z-axis) is a circular patch of light; however, as can be seen from Fig. 4.4a, on a plane AB the circular patch has the least diameter. This is called the circle of least confusion (see Plate 2). It may be mentioned that for an object lying on the axis of a cylindrically symmetric system (like a system of coaxial lenses), the image will suffer only from spherical aberration. All other off-axis aberrations like coma, astigmatism, etc. will be absent.

To see how the rays hitting the refracting surface at different heights could focus to different points on the axis, let us consider the simple case of a plane refracting surface as shown in Fig. 4.5. Let the plane of the refracting surface be chosen as the plane $z=0$. Let P be the object point. The z-axis is chosen to be along the normal (PO) from the point P to the surface. The plane $z=0$ separates two media of refractive indices n_1 and n_2 (see Fig. 4.5); in the figure we have assumed $n_2>n_1$. Consider a ray PM incident on the refracting surface (from the object) at a height h as shown in Fig. 4.5. The refracted ray appears to emerge from the point Q. We assume the origin to be at the point O. Let the z-coordinates of the points P and Q be z_0 and z_1 respectively. Obviously, both z_0 and z_1 would be negative quantities and the distances OP and OQ would be $-z_0$ and $-z_1$ respectively (see Fig. 4.5). We have to determine z_1 in terms of z_0. From Snell's law we know that

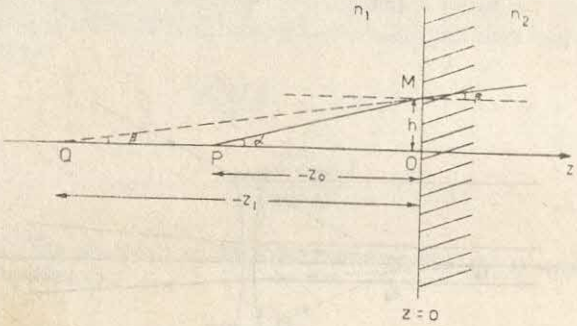


Fig. 4.5 Refraction at a plane surface.

$$\sin \alpha = n \sin \beta \qquad (4.3\text{-}1)$$

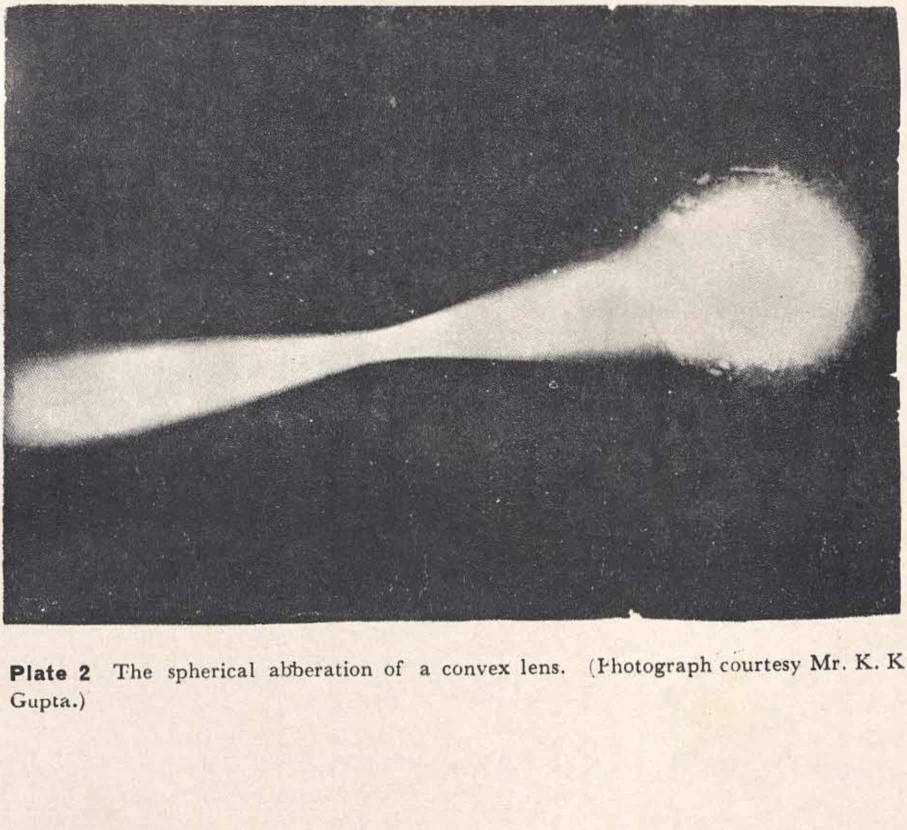


Plate 2 The spherical abberation of a convex lens. (Photograph courtesy Mr. K. K. Gupta.)

where α and β are the angles that the incident and refracted rays make with the z-axis and

$$n = \frac{n_2}{n_1} \tag{4.3-2}$$

Now, from Fig. 4.5 we have

$$(-z_1) = h \cot \beta = \frac{h}{\sin \beta} \sqrt{1-\sin^2 \beta}$$

or

$$z_1 = -\frac{nh}{\sin \alpha} \sqrt{1 - \frac{1}{n^2} \sin^2 \alpha} \tag{4.3-3}$$

where we have used Eq. (4.3-1). Since

$$\sin \alpha = \frac{h}{\sqrt{h^2+z_0^2}} \tag{4.3-4}$$

we obtain

$$z_1 = -\frac{nh}{h} (h^2+z_0^2)^{1/2} \left[1 - \frac{1}{n^2} \frac{h^2}{(h^2+z_0^2)} \right]^{1/2} \tag{4.3-5}$$

or

$$z_1 = -n\,|z_0| \left[1 + \frac{h^2}{z_0^2} \right]^{1/2} \left[1 - \frac{h^2}{n^2 z_0^2} \left(1 + \frac{h^2}{z_0^2} \right)^{-1} \right]^{1/2} \tag{4.3-6}$$

The value of z_1 given in Eq. (4.3-6) is an exact expression in terms of z_0. It can at once be seen that the image distance, z_1, is a complicated function of the height h, at which the ray strikes the refracting surface. In the limit of $h \to 0$, i.e. for paraxial rays, we get

$$z_1 = -n\,|z_0| \tag{4.3-7}$$

which is the expression for the image distance in the paraxial region. To the next order of approximation, assuming $|h/z_0| \ll 1$, we get

$$z \simeq -n\,|z_0| \left[1 + \frac{h^2}{2z_0^2} \right] \left[1 - \frac{h^2}{2n^2 z_0^2} \right]$$

$$\simeq -n\,|z_0| \left[1 + \frac{h^2}{2z_0^2 n^2} (n^2-1) \right] \tag{4.3-8}$$

Thus the aberration is given by

$$\Delta z = -\frac{h^2}{2n\,|z_0|} (n^2-1) \tag{4.3-9}$$

The above expression gives the longitudinal spherical aberration. The negative sign implies that the non-paraxial rays appear to emanate from a point which is farther away from the paraxial image point.

From the above example, it can be seen that even a single plane refracting surface suffers from spherical aberration. Thus, spherical refracting surfaces and thin lenses must also suffer from spherical aberration.

The calculation of the spherical aberration even for a single spherical refracting surface is quite cumbersome (see, e.g. Smith, 1928); we just give the final result

$$\Delta z = -\frac{(n_2-n_1)}{2n_2\left(\frac{1}{z_0}+\frac{n_2-n_1}{n_1R}\right)^2}\left(\frac{1}{R}+\frac{1}{z_0}\right)^2\left(-\frac{n_2+n_1}{n_1z_0}+\frac{1}{R}\right)h^2 \qquad (4.3\text{-}10)$$

where R represents the radius of curvature of the surface, n_1 and n_2 represent the refractive indices of the media on the left and right of the spherical surface (see Fig. 4.6). For a plane surface $R=\infty$, Eq. (4.3-10) reduces to Eq. (4.3-9) with $n=n_2/n_1$.

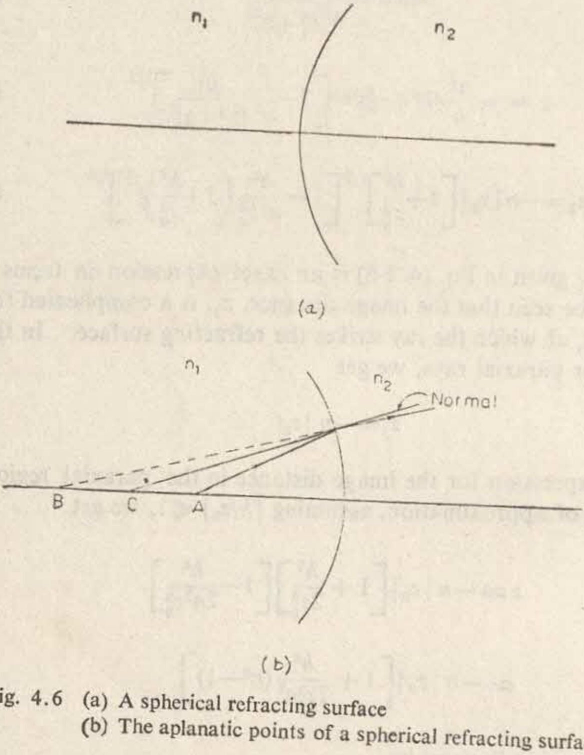


Fig. 4.6 (a) A spherical refracting surface
(b) The aplanatic points of a spherical refracting surface.

Example 4.2 Consider a spherical refracting surface of radius R. Show that for a point A [see Fig. 4.6(b)] such that

$$z_0 = \frac{n_1+n_2}{n_1}R \qquad (4.3\text{-}11)$$

the spherical aberration is zero. Notice that both R and z_0 are negative quantities. The corresponding image point B is at a distance $\frac{n_2-n_1}{n_2}z_0$. The points A and B are known as the aplanatic points and are utilized in microscope objectives.

Solution. For $z_0=\frac{n_1+n_2}{n_1}R$, one of the factors in Eq. (4.3-10) vanishes and the spherical aberration is zero. Indeed, it can be rigorously shown that all rays emanating from the point A appear to diverge from the point B.

Example 4.3 Consider a refracting surface obtained by revolving an ellipse about its major axis. Show that all the rays parallel to the major axis will focus at one of the foci if the eccentricity of the ellipse is equal to n_1/n_2.

Hint: The eccentricity of the ellipse is given by

$$\epsilon=\frac{OF}{a}=\sqrt{1-\frac{b^2}{a^2}}$$

where a and b are the semi-major and semi-minor axes respectively (see Fig. 4.7). If we assume $n_1(QP)+n_2(PF)=n_2(BF)$, then show that the coordinates of the point $P(x, y)$ will satisfy the equation of the ellipse.

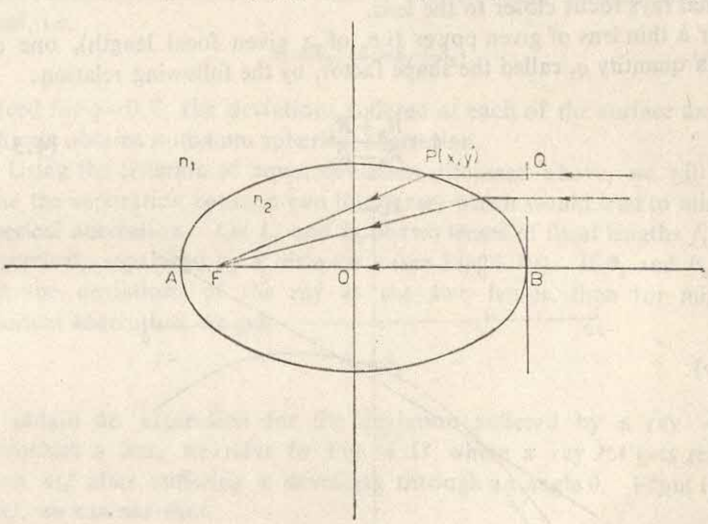


Fig. 4.7 For Example 4.3.

In a similar manner, for a set of rays incident parallel to the axis, one can show that the coefficient of spherical aberration of a thin lens made of a material of refractive index n and placed in air, with the surfaces having radii of curvatures R_1 and R_2, would be given by

$$A=-\frac{f(n-1)}{2n^2}\left[-\left(\frac{1}{R_2}-P\right)^2\left\{\frac{1}{R_2}-P(n+1)\right\}+\frac{1}{R_1^3}\right] \qquad (4.3\text{-}12)$$

where

$$P=\frac{1}{f}=\left[(n-1)\left(\frac{1}{R_1}-\frac{1}{R_2}\right)\right]^{-1} \tag{4.3-13}$$

represents the power of the lens. The coefficient A is such that when it is multiplied by the cube of the height of the ray at the lens, one obtains the lateral spherical aberration. Thus the lateral spherical aberration for rays hitting the lens at a height h would be

$$S_{\text{lat}}=Ah^3=-\frac{f(n-1)h^3}{2n^2}\left[-\left(\frac{1}{R_2}-P\right)^2\left\{\frac{1}{R_2}-P(n+1)\right\}+\frac{1}{R_1^3}\right] \tag{4.3-14}$$

The longitudinal spherical aberration which corresponds to the difference between the marginal focal length and the paraxial focal length would be given by

$$\begin{aligned} S_{\text{long}} &= Ah^2f \\ &= -\frac{(n-1)f^2h^2}{2n^2}\left[\frac{1}{R_1^3}-\left(\frac{1}{R_2}-\frac{n+1}{f}\right)\left(\frac{1}{R_2}-\frac{1}{f}\right)^2\right] \end{aligned} \tag{4.3-15}$$

For a converging lens, S_{long} will always be negative implying that the marginal rays focus closer to the lens.

For a thin lens of given power (i.e. of a given focal length), one can define a quantity q, called the shape factor, by the following relation:

$$q=\frac{R_2+R_1}{R_2-R_1} \tag{4.3-16}$$

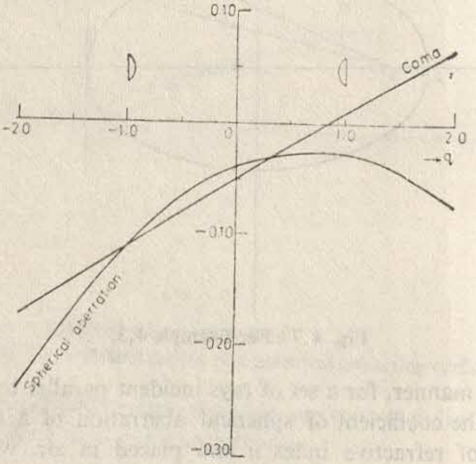


Fig. 4.8 Variation of spherical aberration and coma with the shape factor of a thin lens with $n=1.5$, $f=40$ cm and $h=1$ cm. For calculating the coma we have assumed $\tan\theta=1$, i.e. rays make an angle of 45° with the axis.

where R_1 and R_2 are the radii of curvatures of the two surfaces. For a given focal length of the lens, one can control the spherical aberration by changing the value of q. This procedure is called bending of the lens. Figure 4.8 shows the variation of spherical aberration with q for $n=1.5$, $f=40$ cm (i.e. $P=0.025$ cm^{-1}) and $h=1$ cm. It can be seen that for values of q lying near $q\simeq+0.7$, the (magnitude of the) spherical aberration is minimum (but not zero). Thus, by choosing proper values of the radii, the spherical aberration can be minimized. It may be mentioned that the value $q=+1$ implies $R_2=\infty$ and hence it corresponds to a plano-convex lens with the convex side facing the incident light. On the other hand, for a plano-convex lens with the plane side facing the incident light $R_1=\infty$ and $q=-1$. Thus the spherical aberration is dependent on how the deviation is divided between the surfaces.

The physical reason for the minimum of $|S_{\text{long}}|$ to occur at $q\simeq0.7$ is as follows: It has already been mentioned before that (for a converging lens) the marginal rays undergo a large deviation which results in the spherical aberration (see Fig. 4.4a). As such we should expect the spherical aberration to be minimum when the angle of deviation δ [see Fig. 4.9(a)] is minimum. As in the case of the prism [see Fig. 4.9(b)], this would occur when the deviations suffered at each of the refracting surfacls are exactly equal, i.e.

$$\delta_1=\delta_2;\ (\delta=\delta_1+\delta_2) \tag{4.3-17}$$

Indeed for $q=0.7$, the deviations suffered at each of the surface are equal and one obtains minimum spherical aberration.

Using the criterion of equal deviation discussed above, we will determine the separation between two thin lenses which would lead to minimum spherical aberration. Let L_1 and L_2 be two lenses of focal lengths f_1 and f_2 respectively separated by a distance x (see Fig. 4.10). If θ_1 and θ_2 represent the deviations of the ray at the two lenses, then for minimum spherical aberration, we get

$$\theta_1=\theta_2 \tag{4.3-18}$$

To obtain an expression for the deviation suffered by a ray when it encounters a lens, we refer to Fig. 4.11 where a ray PA gets refracted along AQ after suffering a deviation through an angle θ. From triangle PAQ, we can see that

$$\theta=\theta_1+\theta_2\simeq\frac{h}{v}+\frac{h}{(-u)}=h\left(\frac{1}{v}-\frac{1}{u}\right)=\frac{h}{f} \tag{4.3-19}$$

where we have used the paraxial relation

$$\frac{1}{v}-\frac{1}{u}=\frac{1}{f} \tag{4.3.20}$$

The quantity u is an intrinsically negative quantity. Thus Eq. (4.3-18)

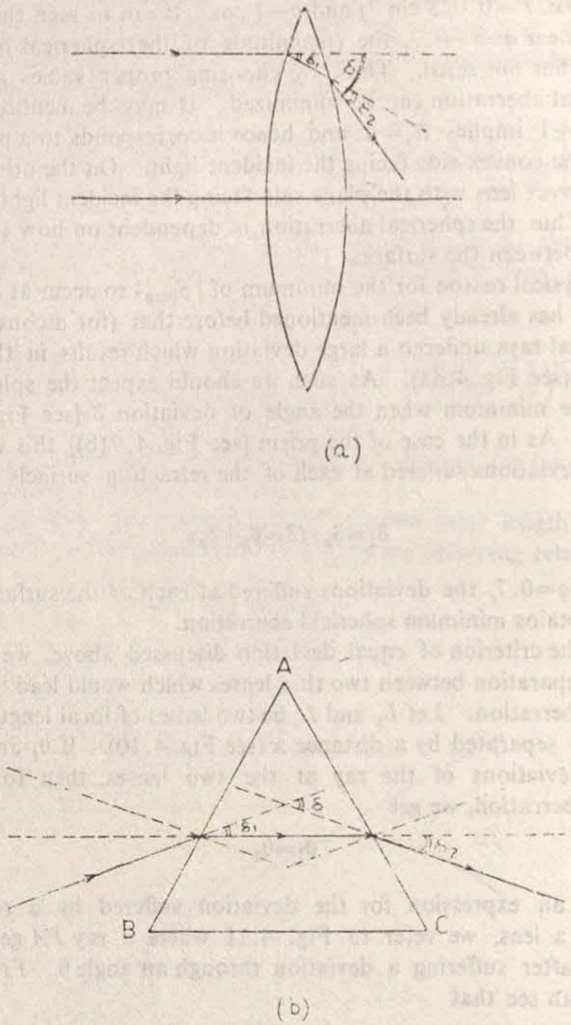


Fig. 4.9 (a) Refraction at the two refracting surfaces of a thin lens; the diagram is exaggerated to show clearly the angles. (b) For a prims, the minimum deviation position corresponds to $\delta_1 = \delta_2$.

becomes

$$\frac{h_1}{f_1}=\frac{h_2}{f_2} \tag{4.3-21}$$

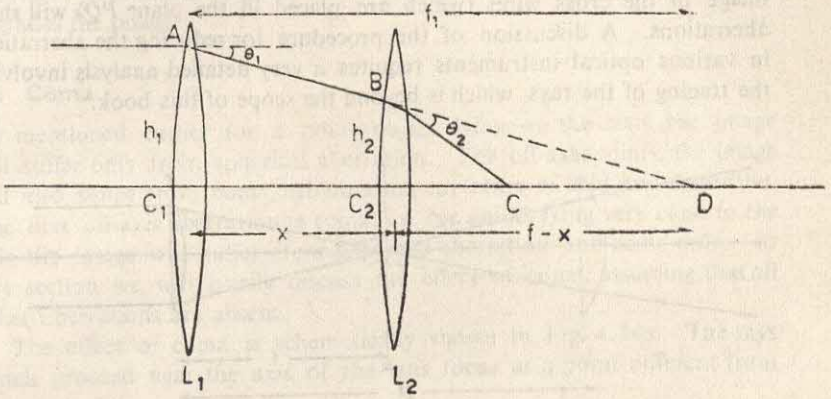


Fig. 4.10 Condition for minimum spherical aberration for a combination of two thin lenses.

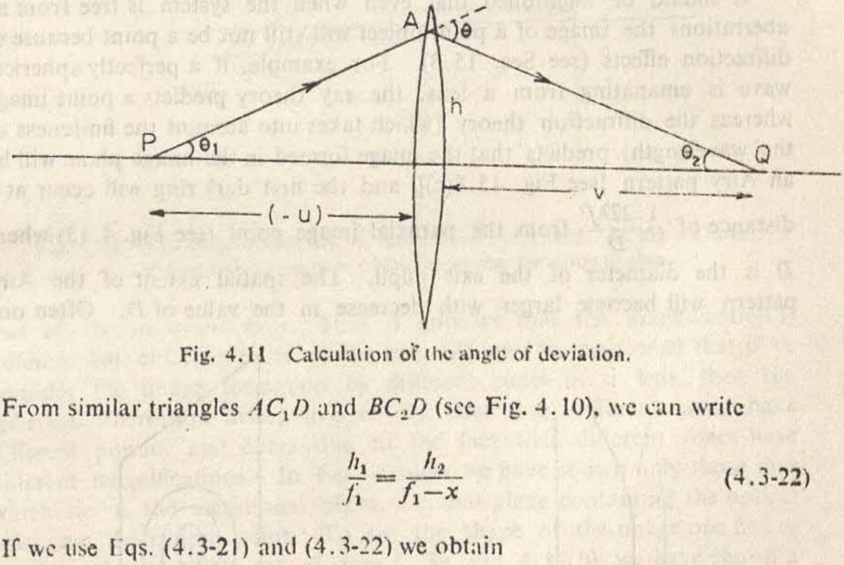


Fig. 4.11 Calculation of the angle of deviation.

From similar triangles AC_1D and BC_2D (see Fig. 4.10), we can write

$$\frac{h_1}{f_1}=\frac{h_2}{f_1-x} \tag{4.3-22}$$

If we use Eqs. (4.3-21) and (4.3-22) we obtain

$$x=f_1-f_2 \tag{4.3-23}$$

Thus the spherical aberration of a combination of two thin lenses is a minimum when their separation is equal to the difference in their focal lengths. Indeed. in the Huygens eyepiece (see Fig. 4.12), the focal length of the field lens is $3f$ where f represents the focal length of the eye lens. The distance between the two lenses is $2f$. We can immediately see that the conditions for achromatism (see Eq. 4.2-15) and minimum spherical aberration (see Eq. 4.3-23) are simultaneously satisfied. Since, the eyepiece as a whole is corrected and the individual lenses are not, the image of the cross wires (which are placed in the plane PQ) will show aberrations. A discussion of the procedure for reducing the aberrations in various optical instruments requires a very detailed analysis involving the tracing of the rays, which is beyond the scope of this book.

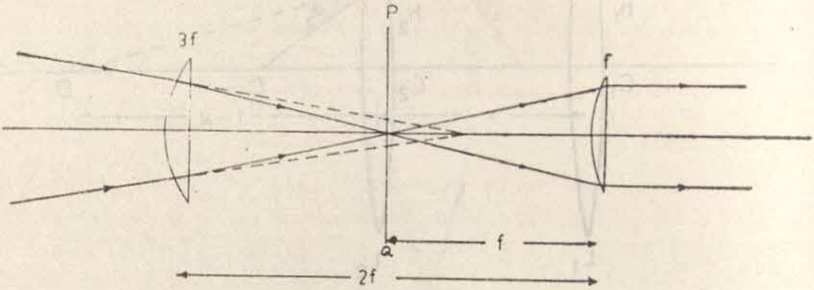


Fig. 4.12 The Huygens eyepiece.

It should be mentioned that even when the system is free from all aberrations the image of a point object will still not be a point because of diffraction effects (see Sec. 15.3). For example, if a perfectly spherical wave is emanating from a lens, the ray theory predicts a point image whereas the diffraction theory (which takes into account the finiteness of the wavelength) predicts that the image formed in the image plane will be an Airy pattern [see Fig. 15.8(c)], and the first dark ring will occur at a distance of $\frac{1.22\lambda f}{D}$ from the paraxial image point (see Fig. 4.13) where D is the diameter of the exit pupil. The spatial extent of the Airy pattern will become larger with decrease in the value of D. Often one

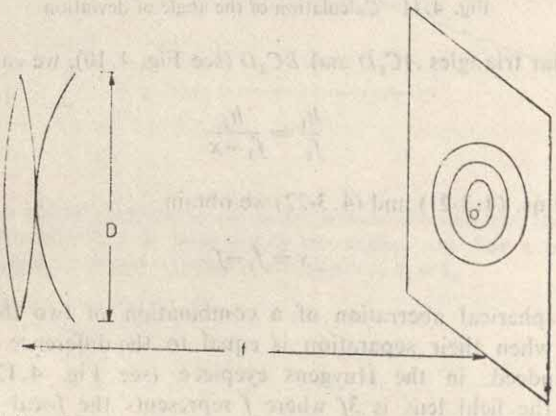


Fig. 4.13 A perfectly spherical wave (converging to the point O) will produce an Airy pattern in the image plane

uses 'stop' to restrict to the paraxial region; however, if the diameter of the 'stop' is made very small then the diffraction effects would predominate. Indeed, a camera gives best image when $f/D \approx 5.6$; at high apertures aberrations degrade the image and at low apertures diffraction degrades the image.

4.4 Coma

As mentioned earlier for a point object lying on the axis the image will suffer only from spherical aberration. For off-axis points, the image will also suffer from coma, astigmatism, curvature of field and distortion. The first off-axis aberration is coma; i.e. for points lying very close to the axis the image will suffer from spherical aberration and coma only. In this section we will briefly discuss the effect of coma, assuming that all other aberrations are absent.

The effect of coma is schematically shown in Fig. 4.14a. The rays which proceed near the axis of the lens focus at a point different from that of the marginal rays. Thus it appears that the magnification is different for different parts of the lens. It may be mentioned that if we consider the image formation by different zones of a lens, then the spherical aberration arises due to the fact that different zones have different powers and coma due to the fact that different zones have different magnifications. In Fig. 4.14(a) we have shown only those rays which lie in the meridional plane, i.e. that plane containing the optical axis and the object point. To see the shape of the image one has to consider the complete set of rays.* In Fig. 4.14(b) we have shown a three-dimensional perspective in which we have considered a set of rays which hit the lens at the same distance from the centre. Rays which

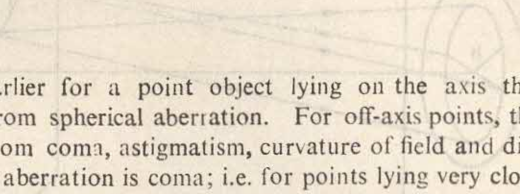
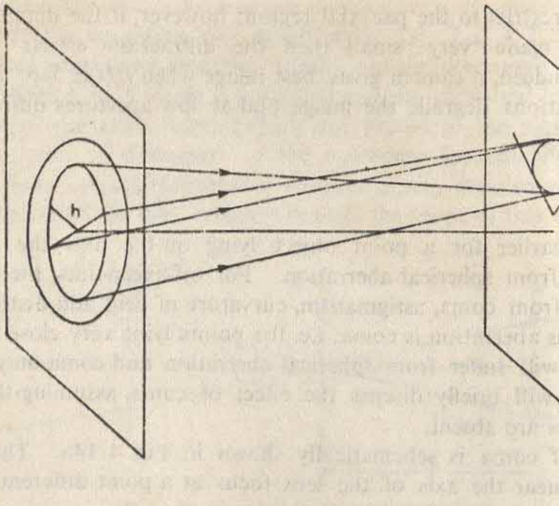


Fig. 4.14 The image formation in the presence of Coma. In (a) we have shown only those rays which lie in the meridional plane.

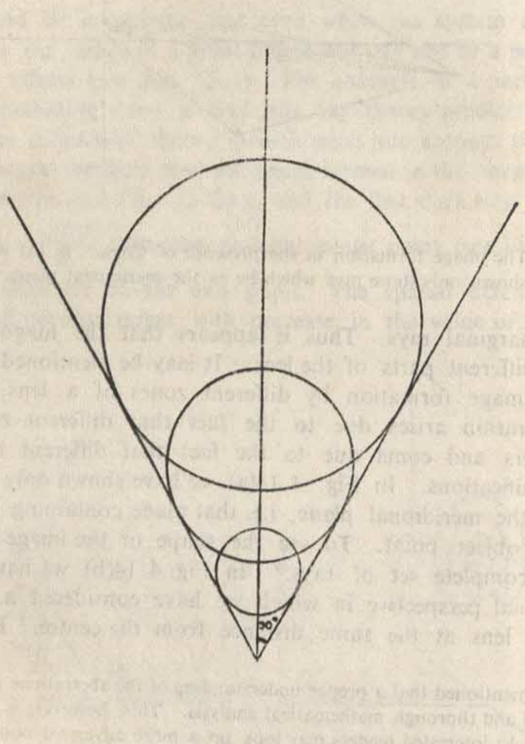
---

*It must be mentioned that a proper understanding of the aberrations can only be had by a careful and thorough mathematical analysis. This, however, is beyond the scope of this book; interested readers may look up a more advanced book like Born and Wolf (1975)



(b)

Fig. 4.14 (b) a three dimensional perspective is shown.



(c)

Fig. 4.14 (c) we have shown the composite image.

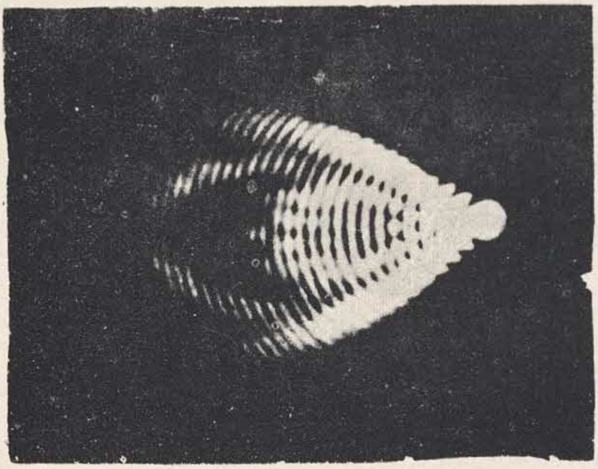


Plate 3 Image of a point source showing coma. (After H. F. Meiners, *Physics Demonstration Experiments*, Vol. II, The Ronald Press Co., New York, 1970; used with permission.)

intersect the lens at diametrically opposite points focus to a single point on the paraxial image plane. These different pairs of rays focus to different points in the image plane such that these foci lie on a circle. The radius of the circle and the distance at which the centre lies from the ideal image point measures the coma. As the radius of the zone [shown as h in Fig. 4.14(b)] increases, the centre of the circle also shifts away from the ideal image. Thus the composite image will have a form as shown in Fig. 4.14(c). The image of a point object thus has a comet like appearance and hence the name coma (see Plate 3).

For a parallel bundle of rays incident on a lens and inclined at an angle θ with the z-axis (see Fig. 4.15), one can show (see, e.g. Born and Wolf, 1975) that the coma in the image is given by

$$\text{Coma} = \frac{3\,(n-1)}{2}\, fh^2 \tan^2\theta \left[\frac{(n-1)(2n+1)}{nR_1R_2} - \frac{n^2-n-1}{n^2R_1^2} - \frac{n}{R_2^2}\right] \quad (4.4\text{-}1)$$

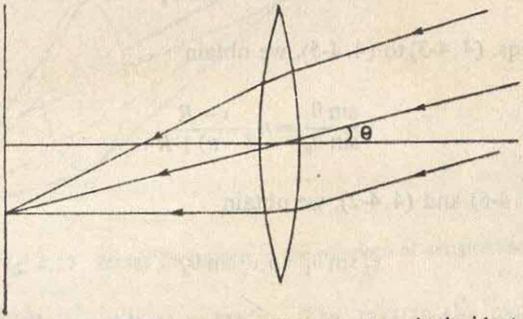


Fig. 4.15 Parallel rays (inclined at an angle θ with the axis) incident on a thin lens.

In Fig. 4.8 we have plotted the variation of coma with the shape factor q. It can immediately be seen that for a lens with $q=+0.8$, coma is zero. It can also be seen that both spherical aberration and coma are close to a minimum for a plano-convex lens (with the convex side facing the incident light) for which $q=1.0$ and as such plano-convex lenses are extensively used in eye-pieces.

We will now derive the condition to be satisfied by an optical system which is free from spherical aberration as well as coma. The relation is called the Abbe-sine condition. We will derive it for the particular case of a spherical refracting surface, but the condition is more general (see e.g. Born and Wolf, 1975). Let SS' be the spherical surface of refraction of radius of curvature R separating two media of relative refractive index n as shown in Fig. 4.16. Let Q be the image of the point P. Let P' be an off axis object point at a distance x_1 from P. We have to determine the condition such that Q' is the image of P'.

From similar triangles $PP'C$ and $QQ'C$, we have

$$\frac{x_1}{x_2} = \frac{(-u)+R}{v-R} \quad (4.4\text{-}2)$$

where u and v represent the object and image distances from the pole of the spherical surface. Since Q is the image of P, a ray PT, after refraction passes through the point Q. Applying Snell's law at the point T, we get

$$\sin \alpha = n \sin \beta \tag{4.4-3}$$

where α and β are shown in Fig. 4.16. Applying the law of sines to the triangle PTC, we obtain

$$\frac{\sin (\pi - \alpha)}{(-u)+R} = \frac{\sin \theta_1}{R}$$

where $\angle TPC = \theta_1$. Thus

$$\frac{(-u)+R}{R} = \frac{\sin \alpha}{\sin \theta_1} \tag{4.4-4}$$

Similarly

$$\frac{v-R}{R} = \frac{\sin \beta}{\sin \theta_2} \tag{4.4-5}$$

Thus using Eqs. (4.4-3) to (4.4-5), we obtain

$$\frac{\sin \theta_1}{\sin \theta_2} = n \, \frac{v-R}{(-u)+R} \tag{4.4-6}$$

From Eqs. (4.4-6) and (4.4-2), we obtain

$$x_1 \sin \theta_1 = n \, x_2 \sin \theta_2 \tag{4.4-7}$$

This is called the Abbe-sine condition. Thus if this condition is satisfied then points near the axis are imaged sharply even for large values of θ_1.

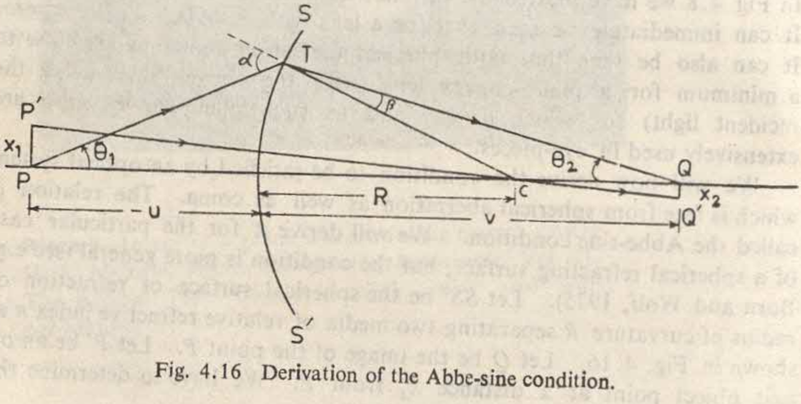


Fig. 4.16 Derivation of the Abbe-sine condition.

4.5 Astigmatism and Curvature of Field

When an optical system is free from spherical aberration and coma then the system will image sharply those object points lying on or near the axis. But for points far away from the axis, the image of a point will not be a point and then the optical system is said to be afflicted with astigmatism.

Consider an object point *P* far away from the axis. The plane containing the axis and the object point is called the meridional plane and the plane perpendicular to the meridional plane (containing the axis) is called the sagittal plane. Fig. 4.17 shows the image formation when the optical system suffers from astigmatism only. The rays in the meridional plane converge at a different point as compared to those in the sagittal plane. For example, the rays *PA* and *PB* focus at the point *T* and the rays *PC* and *PD* focus at a point *S* which is different from *T*. Since at the point *T*, the rays in the sagittal plane have not still focused, one in fact has a focal line which is normal to the meridional plane. This focal line *T* is called the tangential focus. Similarly, since at *S*, the rays in the meridional plane have defocussed, one obtains a focal line lying in the tangential plane; this is called the sagittal focal line. The distance between *S* and *T* is a measure of astigmatism.

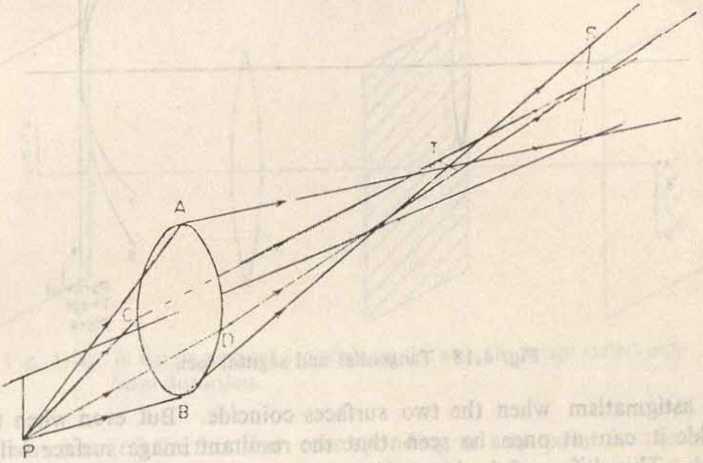


Fig. 4.17 Image formation in the presence of astigmatism.

To see the origin of astigmatism one observes that for a point on the axis (when the lens is free from other aberrations) the wavefront emergin from the lens is spherical and thus as the wavefront progresses, it converges to a single point. But when the object point is non-axial, then the emerging wavefront is not spherical and thus as the wavefront converges, it does not focus to a point but to two lines, which are normal to each other and called the tangential and the sagittal focal lines. Somewhere between the two focal lines, the image is circular in shape and is called the circle of least confusion.

The distance between the tangential and the sagittal foci increases as the object point moves away from the axis. Thus the tangential foci and the sagittal foci of points at different distances from the axis lie on two surfaces as shown in Fig. 4.18. The optical system will be said to be free from astigmatism when the two surfaces coincide. But even when they coincide it can at once be seen that the resultant image surface will be curved. This defect of the image is termed curvature of the field.

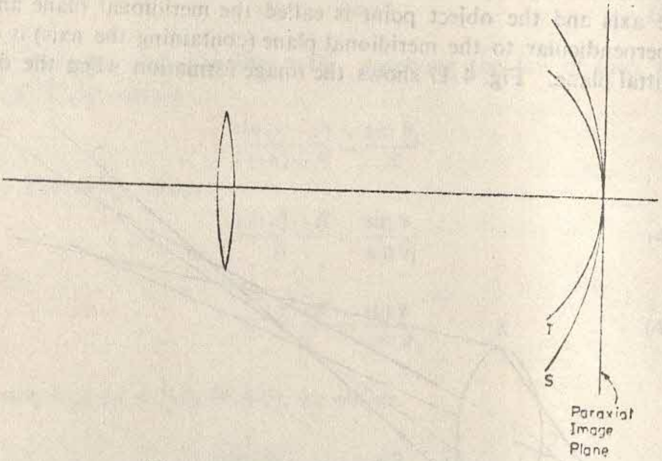
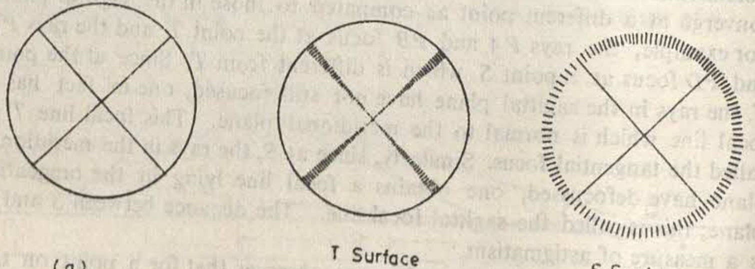


Fig. 4.18 Tangential and sagittal foci.

As an example of image formation in the presence of astigmatism, consider a spoked wheel coaxial with the lens axis as shown in Fig. 4.19(a).



(a)

T Surface
(b)

S Surface
(c)

Fig. 4.19 (a) shows a spoked object coaxial with the axis of the lens; (b) and (c) show the images on the *T*-surface and *S*-surface respectively.

Since on the *T* surface the image of a point source is a line perpendicular to the meridional plane, on the *T*-surface, the complete rim of the wheel will be in focus while the spokes will be out of focus as shown in Fig. 4.19(b). Similarly, since on the *S*-surface the image of a point is a line in the meridional plane, the spokes will be in focus and the rim will not be in focus as shown in Fig. 4.19(c).

4.6 Distortion

The last of the Seidel aberrations is called distortion and is caused by non-uniform magnification of the system. When we discussed spherical aberration we had mentioned that for a point object on the axis of the optical system, the image will suffer only from spherical aberration. Similarly, if we have a pinhole on the axis at any plane of the optical system (see Fig. 4.20), then the image will suffer only from distortion.

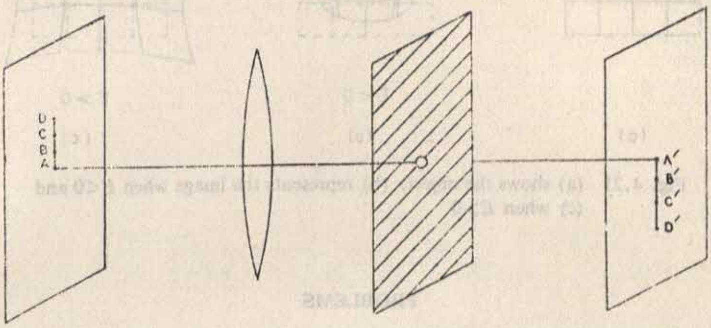


Fig. 4.20 In the presence of a pinhole on the axis, the image suffers only from distortion.

This is because of the fact that corresponding to any point in the object plane, only one of the rays emanating from this point will pass through the pinhole; consequently, all other aberrations will be absent. Obviously, for such a configuration, each point will be imaged as a point but if the system suffers from non uniform magnification, the image will be distorted. This can be illustrated if we consider the imaging of four equally spaced points *A*, *B*, *C*, and *D* which are imaged as *A′*, *B′*, *C′* and *D′* respectively. Mathematical analysis shows that

$$X_d = M x_0 + E(x_0^2 + y_0^2)\, x_0 \tag{4.6-1}$$

and

$$Y_d = M y_0 + E(x_0^2 + y_0^2)\, y_0 \tag{4.6-2}$$

where (x_0, y_0) and (X_d, Y_d) represent the coordinates of the object and the image point respectively, M represents the magnification of the system and E represents the coefficient of distortion. Figure 4.21(b) corresponds to a negative value of E and is known as barrel distortion. The distortion of the image can be easily understood if we consider the imaging of a square grid as shown in Fig. 4.21(a). Assuming unit magnification (i.e. $M=1$), the points having coordinates $(0,0)$, $(h,0)$, $(2h,0)$ $(3h,0)$, $(0,h)$, $(0,2h)$, $(0,3h)$, (h,h), $(h,2h)$, $(2h,h)$, . . . are imaged at $(0,0)$, $(h+Eh^3,0)$, $(2h+8Eh^3,0)$, $(3h+27Eh^3,0)$, $(0,h+Eh^3)$, $(0,2h+8Eh^3)$, $(0,3h+27Eh^3)$, $(h+Eh^3,h+Eh^3)$, $(h+Eh^3,2h+8Eh^3)$, $(2h+8Eh^3,h+Eh^3)$, . . . respectively. If the reader actually plots these points, then for $E<0$, he would obtain a figure like the one shown in Fig. 4.21(b). Similarly for $E>0$, he would obtain Fig. 4.21(c). Notice that each point is imaged at a point, but the image is distorted because of non-uniform magnification.

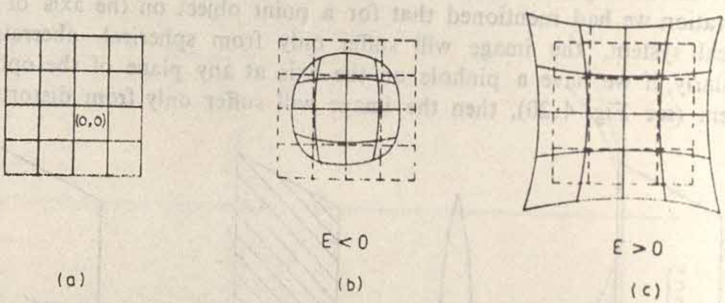


Fig. 4.21 (a) shows the object, (b) represents the image when $E<0$ and (c) when $E>0$.

PROBLEMS

1. Consider a plane glass slab of thickness d made of a material of refractive index n, placed in air. By simple application of Snell's law obtain an expression for the spherical aberration of the slab. What are other kinds of aberrations that the image will suffer from?

   (Ans: Spherical Aberration $= -\frac{(n^2-1)\,dh^2}{2nu^2}$, where h is the height at which the ray strikes the slab, and u is the distance of the object point from the front surface of the slab).
2. Why can't you obtain an expression for the spherical aberration of a plane glass slab from Eq. (4.3-12) by tending R_1, R_2 to ∞?
3. Obtain an expression for the chromatic aberration in the image formed by a plane glass slab.
4. Does the image formed by a plane mirror suffer from any aberration?
5. Calculate the longitudinal spherical aberration of a thin plano-convex lens made of a material of refractive index 1.5 and whose curved surface has a radius of curvature of 10 cm, for rays incident at a height of 1 cm. Compare the values of the aberration when the convex side and the plane side face the incident light.
6. Consider a lens made up of a material of refractive index 1.5 with a focal length 25 cm. Assuming $h=0.5$ cm and $\theta=45°$, obtain the spherical aberration and coma for the lens for various values of the shape factor q and plot the variation in a manner similar to that shown in Fig. 4.8.
7. An achromatic cemented doublet of focal length 25 cm is to be made from a combination of an equiconvex flint glass lens ($n_b=1.50529$, $n_r=1.49776$) and a crown glass lens ($n_b=1.66270$, $n_r=1.64357$). Calculate the radii of curvatures of the different surfaces and the focal lengths of each of the two lenses.

REFERENCES AND SUGGESTED READING

Born, M. and Wolf, E. (1975), *Principles of Optics*, Pergamon Press. Oxford (England).

CAGNET, M., FRANCON, M. and THIERR, J.C. (1962), *Atlas of Optical Phenomena*, Springer-Verlag, Berlin (Germany).

HOPKINS, H.H. (1950), *Wave Theory of Aberrations*, Oxford University Press, London (England).

SMITH, C.J. (1960), A Degree Physics, Part III, *Optics*, Edward Arnold Publishers, London.

WELFORD, W.T. (1962), *Geometrical Optics*, North Holland Publishing Co., Amsterdam (Netherlands).

WELFORD, W.T. (1974), *Aberrations of the Symmetrical Optical System*, Academic Press, New York (USA).

PART II

Vibration and Waves

5 Simple Harmonic Motion

I heard a thousand blended notes...

WORDSWORTH

5.1 Definition

A periodic motion is a motion which repeats itself after regular intervals of time and the simplest kind of periodic motion is a simple harmonic motion in which the displacement varies sinusoidally with time. In order to understand the simple harmonic motion we consider a point P rotating on the circumference of a circle of radius a with an angular velocity ω (see Fig. 5.1). We choose the centre of circle as our origin and we assume

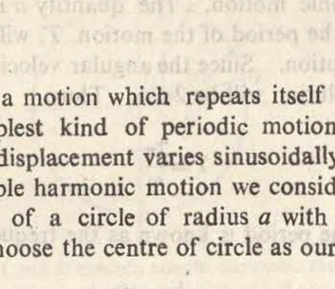


Fig. 5.1 The point P is rotating in the anticlockwise direction on the circumference of a circle of radius a, with uniform angular velocity ω. The foot of the perpendicular on any one of the diameters executes simple harmonic motion. P_0 is the position of the point at $t = 0$.

that at $t=0$ the point P lies on the x-axis (i.e. at the point P_0). At an arbitrary time t the point will be at the position P where $\angle POX=\omega t$.

Let A be the foot of the perpendicular from the point P on the x-axis. Clearly, the distance

$$OA = a\cos\omega t, \tag{5.1-1}$$

and as the point P rotates on the circumference of the circle, the point A moves to and fro about the origin on the diameter. When the point P is at P_1, then the foot of the perpendicular is at O. This can also be seen from Eq. (5.1-1) because when P coincides with P_1, $\omega t=\pi/2$ and hence $a\cos\omega t=a\cos\pi/2=0$. As the point still moves further, the foot of the perpendicular would lie on the other side of the origin and thus OA would be negative as is also evident from Eq. (5.1-1) because ωt would then be greater than $\pi/2$. When P coincides with P_2, then $OA=OP_2=-a$. When the point P moves from P_2 to P_3, OA starts decreasing and it finally goes to zero when P coincides with P_3. After P crosses P_3, OA starts increasing again and finally acquires the value a when P coincides with P_0. After crossing the point P_0, the motion repeats itself.

A motion in which the displacement varies sinusoidally with time [as in Eq. (5.1-1)] is known as a simple harmonic motion. Thus, when a point rotates on the circumference of a circle with a uniform angular velocity, the foot of the perpendicular on any one of its diameters will execute simple harmonic motion. The quantity a is called the amplitude of the motion, and the period of the motion, T, will be the time required to complete one revolution. Since the angular velocity is ω, the time taken for one complete revolution will be $2\pi/\omega$. Thus

$$T=\frac{2\pi}{\omega} \tag{5.1-2}$$

The inverse of the time period is known as the frequency.

$$\nu=\frac{1}{T}=\frac{\omega}{2\pi}$$

or

$$\omega=2\pi\nu \tag{5.1-3}$$

It should be pointed out that we could as well have studied the motion of the point B, which is the foot of the perpendicular from the point P on the y-axis. The distance OB is given by

$$OB=y=a\sin\omega t. \tag{5.1-4}$$

We had conveniently chosen $t=0$ as the time when P was on the x-axis. The choice of the time $t=0$ is arbitrary and we could have chosen time $t=0$ to be the instant when P was at P' (see Fig. 5.2). If the angle $\angle P'OX=\theta$ then the projection on the x-axis at any time t would be given by

$$OA=x=a\cos(\omega t+\theta) \tag{5.1-5}$$

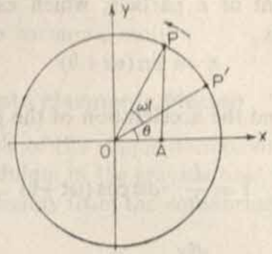


Fig. 5.2 At $t=0$, the point P is at P' and, therefore, the initial phase is θ.

The quantity $(\omega t+\theta)$ is known as the phase of the motion and θ represents the initial phase. It is obvious from the above discussion that the value of θ is quite arbitrary and depends on the instant from which we start measuring time.

We next consider two points P and Q rotating on the circle with the same angular velocity and P' and Q' be their respective positions at $t=0$. Let the angles $\angle P'OX$ and $\angle Q'OX$ be θ and φ respectively (see Fig. 5.3).

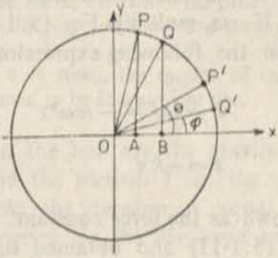


Fig. 5.3 The points A and B execute simple harmonic motions with the same frequency ω. The initial phases of A and B are θ and φ respectively.

Clearly at an arbitrary time t, the distance of the foot of the perpendiculars from the origin would be

$$x_P = a\cos(\omega t+\theta) \tag{5.1-6a}$$

$$x_Q = a\cos(\omega t+\varphi) \tag{5.1-6b}$$

The quantity

$$(\omega t+\theta)-(\omega t+\varphi) = \theta-\varphi \tag{5 1-7}$$

represents the phase difference between the two simple harmonic motions and if $\theta-\varphi=0$ (or an even multiple of π) the motions are said to be in phase, and if $\theta-\varphi=\pi$ (or an odd multiple of π) the motions are said to be out of phase. If we choose a different origin of time, the quantities θ and φ would change by the same additive constant; consequently, the phase difference $(\theta-\varphi)$ is independent of the choice of the instant $t=0$.

Thus the displacement of a particle, which executes simple harmonic motion, can be written as

$$x = a\sin(\omega t+\theta) \tag{5.1-8}$$

Therefore, the velocity and the acceleration of the particle would be given by the following equations

$$v = \frac{dx}{dt} = a\omega\cos(\omega t+\theta) \tag{5.1-9}$$

and

$$f = \frac{d^2x}{dt^2} = -a\omega^2\sin(\omega t+\theta)$$

$$= -\omega^2 x \tag{5.1-10}$$

Eq. (5.1-10) shows that the acceleration of the particle is proportional to the displacement and the negative sign indicates that the acceleration is always directed towards the origin. Eq. (5.1-10) can be used to define the simple harmonic motion as the motion of a particle in a straight line in which the acceleration is proportional to the displacement from a fixed point (on the straight line) and always directed towards the fixed point. (Here the point $x = 0$ is the fixed point and usually referred to as the equilibrium position). If we multiply Eq. (5.1-10) by the mass of the particle, then we obtain the following expression for the force acting on the particle

$$F = mf = -m\omega^2 x$$

or

$$F = -kx \tag{5.1-11}$$

where $k\ (=m\omega^2)$ is known as the force constant. We could have equally well started from Eq. (5.1-11) and obtained simple harmonic motion. This could easily be seen by noting that since the force is acting in the x direction, the equation of motion would be

$$m\frac{d^2x}{dt^2} = F = -kx$$

or

$$\frac{d^2x}{dt^2} + \frac{k}{m}x = 0$$

or

$$\frac{d^2x}{dt^2} + \omega^2 x = 0 \tag{5.1-12}$$

where $\omega^2 = k/m$. The general solution of Eq. (5.1-12) can be written in the form

$$x = A\sin\omega t + B\cos\omega t \tag{5.1-13}$$

which can be rewritten in either of the following forms

$$x = a\sin(\omega t+\theta) \tag{5.1-14}$$

or
$$x = a\cos(\omega t + \theta) \tag{5.1-15}$$
which describes a simple harmonic motion.

5.2 Examples of Simple Harmonic Motion

(i) The simplest example of the simple harmonic motion is the motion of the bob of a simple pendulum in the gravitational field. If the bob of the pendulum is displaced slightly from the equilibrium position (see Fig. 5.4)

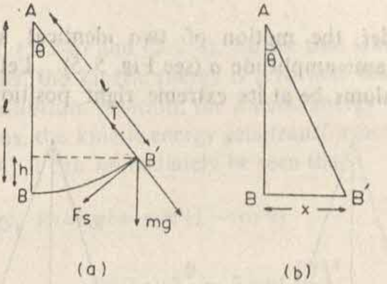


Fig. 5.4 (a) The forces acting on the bob of the pendulum when it is displaced from its equilibrium position. The restoring force is F_s which is equal to $mg\sin\theta$.

(b) If the angle θ is small, the motion of the bob can be approximately assumed to be in a straight line.

then the forces acting on the bob are the gravitational force mg acting vertically downwards and the tension T, in the direction $B'A$. In the equilibrium position (AB) the tension is equal and opposite to the gravitational force. However, in the displaced position the tension T is not in the direction of the gravitational force and if we resolve the gravitational force along the direction of the string and perpendicular to it, we see that the component $mg\cos\theta$ balances the tension in the string and the component $mg\sin\theta$ is the restoring force. The motion of the bob is along the arc of a circle but if the length of the pendulum is large and the angle θ is small, the motion can be assumed to be approximately in a straight line (see Fig. 5.4b). Under such an approximation we may assume that the force is always directed towards the point B and the magnitude of this force will be

$$mg\sin\theta \simeq mg\frac{x}{l} \tag{5.2-1}$$

Thus the equation of motion will be

$$F = m\frac{d^2x}{dt^2} = -mg\,\frac{x}{l} \tag{5.2-2}$$

or

$$\frac{d^2x}{dt^2} + \omega^2 x = 0 \tag{5.2-3}$$

where $\omega^2 = g/l$. Equation (5.2-3) is of the same form as Eq. (5.1-12); thus the motion of the bob is simple harmonic with its time period given by the following equation

$$T = \frac{2\pi}{\omega} = 2\pi\sqrt{\frac{l}{g}} \tag{5.2-4}$$

It should be pointed that the expression for the time period is fairly accurate (i.e. the motion is approximately simple harmonic) as long as $0 \lesssim 4^\circ$.

We next consider the motion of two identical simple pendulums vibrating with the same amplitude a (see Fig. 5.5). Let, at $t=0$, the bob of one of the pendulums be at its extreme right position (Fig. 5.5a) and the bob of the other pendulum be at the equilibrium position, moving towards the right (Fig. 5.5b). If we measure the displacement from the equilibrium positions of the pendulums, then the displacements would be given by

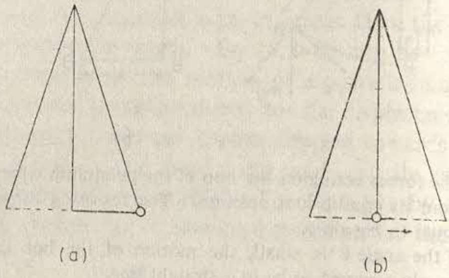


Fig. 5.5 (a) and (b) show the motion of two identical pendulums which are vibrating with the same amplitude but having a phase difference of $\pi/2$. The small circles denote the position of the bobs at $t=0$.

$$x_1 = a\cos\omega t$$

$$x_2 = a\sin\omega t = a\cos\left(\omega t - \frac{\pi}{2}\right) \tag{5.2-5}$$

Thus the two bobs execute simple harmonic motion with a phase difference of $\pi/2$ and in fact the first pendulum is ahead in phase by $\pi/2$. It may be mentioned that in Fig. 5.5(b) if the bob was moving towards the left, then the equation of motion would have been

$$x_2 = -a\sin\omega t = a\cos(\omega t + \pi/2)$$

and then the second pendulum would have been ahead in phase by $\pi/2$.

Since, in general, the displacement of the bob of the pendulum can be written as

$$x = a\cos(\omega t+\varphi) \qquad (5.2\text{-}6)$$

the velocity of the particle would be given by

$$\frac{dx}{dt} = -a\omega\sin(\omega t+\varphi) \qquad (5.2\text{-}7)$$

Thus the kinetic energy of the mass would be

$$T = \tfrac{1}{2}m\left(\frac{dx}{dt}\right)^2 = \tfrac{1}{2}ma^2\omega^2\sin^2(\omega t+\varphi) \qquad (5.2\text{-}8)$$

Comparing Eqs. (5.2-6) and (5.2-8), we see that when the particle is at its extreme positions, the kinetic energy is zero and when the particle passes through the equilibrium position, the kinetic energy is maximum. At the extreme positions, the kinetic energy gets transformed into potential energy. From Fig. 5.4(a) it can immediately be seen that

Potential Energy, $V = mgh = mgl(1-\cos\theta)$

$$= mgl\,2\sin^2\frac{\theta}{2} \simeq 2\,mgl\left(\frac{\theta}{2}\right)^2$$

$$\simeq \tfrac{1}{2}mgl\left(\frac{x}{l}\right)^2 = \tfrac{1}{2}m\left(\frac{g}{l}\right)x^2$$

$$= \tfrac{1}{2}m\omega^2x^2 \qquad (5.2\text{-}9)$$

or* $$V = \tfrac{1}{2}m\omega^2a^2\cos^2(\omega t+\varphi) \qquad (5.2\text{-}10)$$

where we have used the fact that $\omega^2 = g/l$. Thus the total energy E would be given by

$$E = T+V$$

$$= \tfrac{1}{2}m\omega^2a^2 \qquad (5.2\text{-}11)$$

which, as expected, is independent of time. We can also see from Eq. (5.2-11) that the energy associated with the simple harmonic motion is proportional to the square of the amplitude and the square of the frequency.

(ii) Another simple example is the motion of a mass m, held by two stretched springs on a smooth table as shown in Fig. 5.6. The two

---

*The expression for potential energy could have been directly written down by noting the fact that if the potential energies at x and at $x+dx$ are V and $V+dV$, then $dV = -F dx = +kx\,dx$. Thus

$$V = \int_0^x kx\,dx = \tfrac{1}{2}kx^2$$

where we have assumed the zero of the potential energy to be at $x=0$.

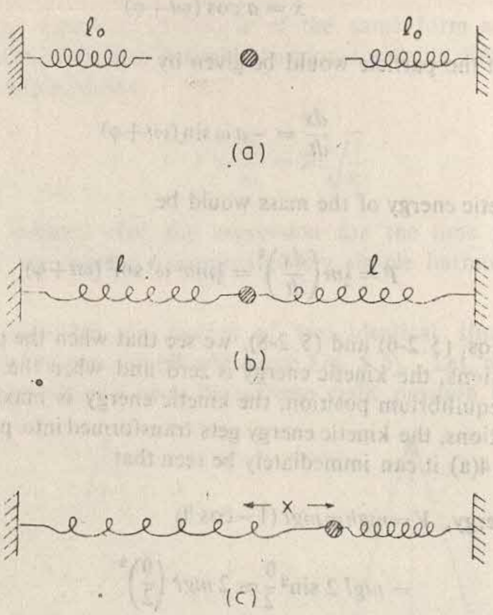


Fig. 5.6 Two springs of natural length l_0 [see (a)] are stretched to a length l [see (b)] to hold the mass. If the mass is displaced by a small distance x from its equilibrium position [see (c)], the mass will execute simple harmonic motion.

springs are of natural length l_0 (Fig. 5.6a) and corresponding to the equilibrium position of the mass, the lengths of the stretched springs are l. If the mass is displaced slightly from the equilibrium position, then the resultant force acting on the mass will be

$$F=k\,[(l-x)-l_0]-k\,[(l+x)-l_0]$$
$$=-2kx \qquad (5.2\text{-}12)$$

where k represents the force constant of the spring. Once again we get a force which is proportional to the displacement and directed towards the equilibrium position and consequently, the motion of the mass on the frictionless table will be simple harmonic in nature.

(iii) When a stretched string (as in a sonometer) is made to vibrate in its fundamental mode (see Fig. 5.7), then each point on the string executes simple harmonic motion with different amplitudes but having the same initial phase. The displacement can be written in the form

$$y=a\sin\left(\frac{\pi}{L}x\right)\cos\omega t.$$

The amplitude is therefore zero at $x=0$ and at $x=L$ and is maximum at $x=L/2$.

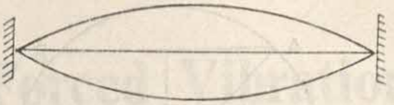


Fig. 5.7 When a string clamped at both the ends is made to vibrate in its fundamental mode, all particles execute simple harmonic motions with same frequency and same initial phase but having different amplitudes.

On the other hand, if the string is vibrating in its first harmonic, then each point on the first half of the string vibrates out of phase with each point on the other half.

PROBLEMS

1. The displacement in a string is given by the following equation

   $$y(x,t)=a\cos\left(\frac{2\pi}{\lambda}x-2\pi\nu t\right)$$

   where a, λ and ν represent the amplitude, wavelength and the frequency of the wave. Assume $a=0.1$ cm, $\lambda=4$ cm, $\nu=1\ \text{sec}^{-1}$. Plot the time dependence of the displacement at $x=0$, 0.5 cm, 1.0 cm, 1.5 cm, 2 cm, 3 cm and 4 cm. Interpret the plots physically.
2. The displacement associated with a standing wave on a sonometer is given by the following equation

   $$y(x,t)=2a\sin\left(\frac{2\pi}{\lambda}x\right)\cos 2\pi\nu t$$

   If the length of the string is L then the allowed values of λ are $2L$, $2L/2$, $2L/3$, ...(see Sec. 10.2). Consider the case when $\lambda=\frac{2L}{5}$, study the time variation of displacement in each loop and show that alternate loops vibrate in phase (with different points in a loop having different amplitudes) and adjacent loops vibrate out of phase.
3. A tunnel is dug through the earth as shown in Fig. 5.8. A mass is dropped at the point A along the tunnel. Show that it will execute simple harmonic motion. What will the time period be?
4. A 1 g mass is suspended from a vertical spring. It executes simple harmonic motion with period 1 sec. By how much distance had the spring stretched when the mass was attached?
5. A stretched string is given simultaneous displacement in the x-and y-directions such that

   $$x(z,t)=a\cos\left(\frac{2\pi}{\lambda}z-2\pi\nu t\right)$$

   and

   $$y(z,t)=a\cos\left(\frac{2\pi}{\lambda}z-2\pi\nu t\right)$$

   Study the resultant displacement (at a particular value of z) as a function of time.

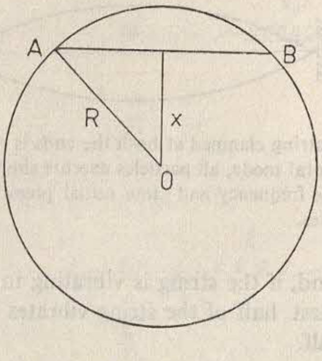


Fig. 5.8 For Problem 3.

6. In the above problem, if

$$x(z,t)=a\cos\left(\frac{2\pi}{\lambda}z-2\pi\nu t\right)$$

and

$$y(z,t)=a\sin\left(\frac{2\pi}{\lambda}z-2\pi\nu t\right),$$

what will be the resultant displacement?

REFERENCES AND SUGGESTED READING

See at the end of Chapter 7

6 Forced Vibrations and Origin of Refractive Index

The souls by nature pitched too high,
By suffering plung'd too low.

GEORGE ELIOT

If a periodic force acts on a vibrating system, the system undergoes what are known as forced vibrations. In this chapter, we will study such kind of vibrations; however, before we do so it is necessary to discuss damped simple harmonic motion.

6.1. Damped Simple Harmonic Motion

In Chapter 5 we had shown that for a particle executing *SMH*, the equation of motion will be of the form

$$\frac{d^2x}{dt^2}+\omega_0^2\, x\,(t)=0 \qquad (6.1\text{-}1)$$

the solution of which is given by

$$x\,(t)=A\cos\,(\omega_0 t+\theta), \qquad (6.1\text{-}2)$$

where A represents the amplitude and ω_0 the angular frequency of motion. Equation (6.1-2) tells us that the motion will continue forever; however, we know that in actual practice the amplitude of any vibrating system (like that of a tuning fork) keeps on decreasing and eventually the system stops vibrating. Similarly, the bob of a pendulum comes to rest after a certain period of time. This phenomenon is due to the presence of damping (or frictional) forces which come into play when the particle is in motion. For a vibrating pendulum, the frictional forces are primarily due to the viscosity of the surrounding medium; consequently, in liquids the frictional forces will be much larger than in gases. In general, the exact dependence of the frictional forces on the velocity of the particle is quite complicated; however, as a first approximation we may assume it to be proportional to the velocity of the particle. This is also consistent with the fact that there are no frictional forces acting on the particle when it is at rest. In this model, the equation of motion will be given by*

$$m\frac{d^2x}{dt^2} = -\Gamma\frac{dx}{dt} - k_0x \tag{6.1-3}$$

where the constant Γ determines the strength of the damping force. Eq. (6.1-3) can be rewritten in the form

$$\frac{d^2x}{dt^2} + 2K\frac{dx}{dt} + \omega_0^2 x(t) = 0 \tag{6.1-4}$$

where

$$2K \equiv \frac{\Gamma}{m} \text{ and } \omega_0 = \sqrt{k_0/m} \tag{6.1-5}$$

In order to solve Eq. (6.1-4) we introduce a new variable $\xi(t)$ which is defined by the following equation:

$$x(t) = \xi(t)e^{-Kt} \tag{6.1-6}$$

Thus

$$\frac{dx}{dt} = \left[\frac{d\xi}{dt} - K\xi(t)\right]e^{-Kt}$$

and

$$\frac{d^2x}{dt^2} = \left[\frac{d^2\xi}{dt^2} - 2K\frac{d\xi}{dt} + K^2\xi(t)\right]e^{-Kt}$$

On substitution in Eq. (6.1-4) we get

$$\frac{d^2\xi}{dt^2} + (\omega_0^2 - K^2)\,\xi(t) = 0. \tag{6.1-7}$$

The above equation is similar to Eq. (6.1-1); however depending on the strength of the damping force, the quantity $(\omega_0^2 - K^2)$ can be positive, negative or zero. Consequently, we must consider three cases.

Case I $(\omega_0^2 > K^2)$

If the damping is small, ω_0^2 is greater than K^2, and the solution of Eq. (6.1-7) would be of the form

---

*We are using the symbol k_0 for the force constant to avoid confusion with the propagation constant k.

$$\xi(t)=A\cos\left[\sqrt{\omega_0^2-K^2}\,t+\theta\right] \tag{6.1-8}$$

or

$$x(t)=Ae^{-Kt}\cos\left[\sqrt{\omega_0^2-K^2}\,t+\theta\right] \tag{6.1-9}$$

where A and θ are constants which are determined from the amplitude and phase of the motion at $t=0$. Eq. (6.1-9) represents a damped simple harmonic motion (see Fig. 6.1). Notice that the amplitude decreases exponentially with time and the time period of vibration $\left(=\dfrac{2\pi}{\sqrt{\omega_0^2-K^2}}\right)$ is greater than in the absence of damping.

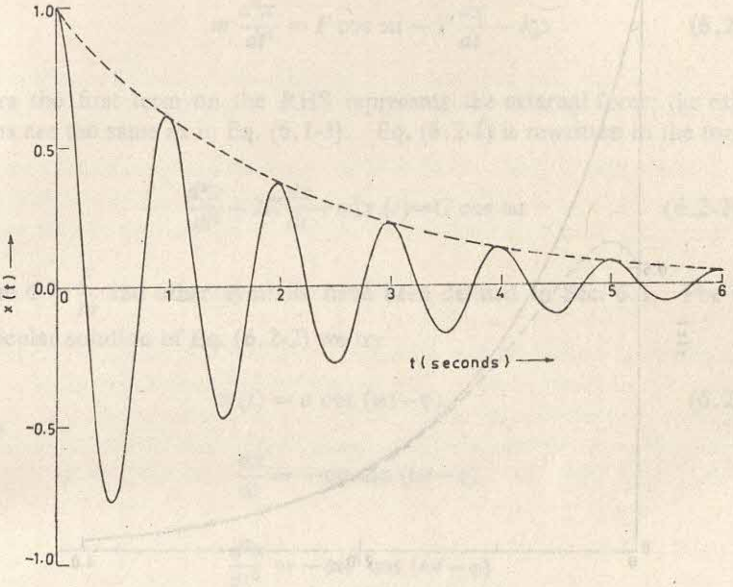


Fig. 6.1 The exponential decrease of amplitude in a damped simple harmonic motion. The figure corresponds to $\dfrac{2\pi}{\sqrt{\omega_0^2-K^2}}=1$ sec and $K=0.5\ \text{sec}^{-1}$.

Case II $(K^2>\omega_0^2)$

If the damping is too large, K^2 is greater than ω_0^2 and Eq. (6.1-7) should be written in the form

$$\frac{d^2\xi}{dt^2}-(K^2-\omega_0^2)\,\xi(t)=0, \tag{6.1-10}$$

the solution of which is given by

$$\xi(t)=A\exp\left[+(K^2-\omega_0^2)^{1/2}t\right]+B\exp\left[-(K^2-\omega_0^2)^{1/2}t\right] \tag{6.1-11}$$

Thus

$$x(t)=A\exp\left(\left[-K+\sqrt{K^2-\omega_0^2}\right]t\right)+B\exp\left(\left[-K-\sqrt{K^2-\omega_0^2}\right]t\right) \quad (6.1\text{-}12)$$

and we can have two kinds of motion; one in which the displacement decreases uniformly to zero or, the other, in which the displacement first increases, reaches a maximum and then decreases to zero (see Fig. 6.2). In either case there are no oscillations and the motion is said to be 'overdamped' or 'dead beat'. A typical example is the motion of a simple pendulum in a highly viscous liquid (like glycerine) where the pendulum can hardly complete a fraction of the vibration before coming to rest.

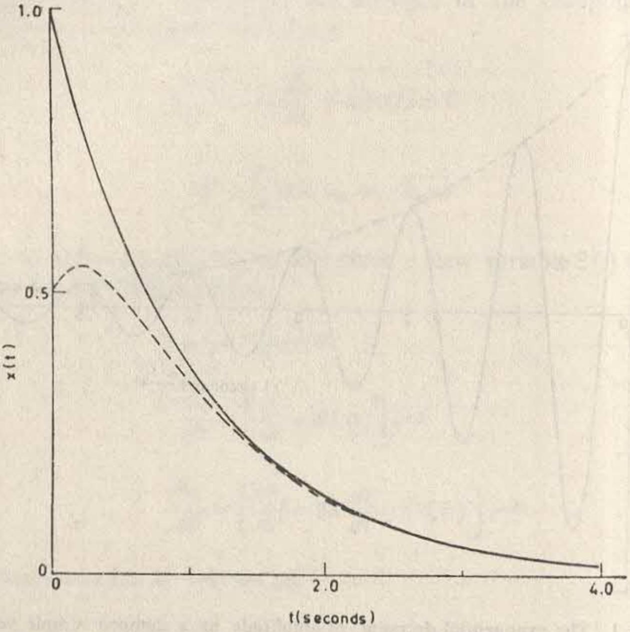


Fig. 6.2 The variation of displacement with time in an overdamped motion. The solid and the dashed curves correspond to $B = 0$ and $B = -A/2$ respectively (see Eq. 6.1-12). In carrying out the calculations we have assumed $K = 2\ \text{sec}^{-1}$ and $\sqrt{K^2-\omega_0^2} = 1\ \text{sec}^{-1}$.

Case III ($K^2 = \omega_0^2$)

When $K^2 = \omega_0^2$, Eq. (6.1-7) becomes

$$\frac{d^2\xi}{dt^2} = 0, \quad (6.1\text{-}13)$$

the solution of which is given by

$$\xi = At + B \quad (6.1\text{-}14)$$

Thus

$$x(t) = (At+B)\,e^{-Kt} \tag{6.1-15}$$

The motion is again non-oscillatory and is said to correspond to 'critical damping'.

6.2 Forced Vibrations

We will now consider the effect of a periodic sinusoidal force* on the motion of a vibrating system. If the frequency of the external force is ω then the equation of motion would be [cf. Eq. (6.1-3)]:

$$m\frac{d^2x}{dt^2} = F\cos\omega t - \Gamma\frac{dx}{dt} - k_0 x \tag{6.2-1}$$

where the first term on the RHS represents the external force; the other terms are the same as in Eq. (6.1-3). Eq. (6.2-1) is rewritten in the form

$$\frac{d^2x}{dt^2}+2K\frac{dx}{dt}+\omega_0^2 x(t)=G\cos\omega t \qquad (6.2\text{-}2)^{**}$$

where $G=\dfrac{F}{m}$ and other symbols have been defined in Sec. 6.1. For the particular solution of Eq. (6.2-2) we try

$$x(t) = a\cos(\omega t-\varphi). \tag{6.2-3}$$

Thus

$$\frac{dx}{dt} = -a\omega\sin(\omega t-\varphi)$$

and

$$\frac{d^2x}{dt^2} = -a\omega^2\cos(\omega t-\varphi)$$

Substituting the above forms for $x(t)$, $\dfrac{dx}{dt}$ and $\dfrac{d^2x}{dt^2}$ in Eq. (6.2-2) we obtain

$$-a\omega^2\cos(\omega t-\varphi)-2Ka\omega\sin(\omega t-\varphi) \\ +a\omega_0^2\cos(\omega t-\varphi)=G\cos[(\omega t-\varphi)+\varphi] \tag{6.2-4}$$

where we have written $G\cos\omega t$ as $G\cos[\omega t-\varphi+\varphi]$. Thus

$$a(\omega_0^2-\omega^2)\cos(\omega t-\varphi)-2Ka\omega\sin(\omega t-\varphi) \\ =G\cos(\omega t-\varphi)\cos\varphi-G\sin(\omega t-\varphi)\sin\varphi \tag{6.2-5}$$

---

*See also Sec. 7.3.

**Notice that the RHS of Eq. (6.2-2) is independent of x; such an equation is said to be inhomogeneous. An equation of the type (6.1-3) is said to be homogeneous.

For Eq. (6.2-5) to be valid for all values of time we must have

$$a(\omega_0^2-\omega^2)=G\cos\varphi \tag{6.2-6}$$

$$2Ka\omega=G\sin\varphi \tag{6.2-7}$$

If we square and add, we get

$$a=\frac{G}{[(\omega_0^2-\omega^2)^2+4K^2\omega^2]^{1/2}} \tag{6.2-8}$$

Further

$$\tan\varphi=\frac{2K\omega}{(\omega_0^2-\omega^2)} \tag{6.2-9}$$

Since K, ω and a are positive, φ is uniquely determined by noting that $\sin\varphi$ should be positive; i.e. φ must be either in the first or in the second quadrant.

To the solution given by Eq. (6.2-3), we must add the solution of the homogeneous equation (Eq. 6.1-4). Thus, assuming ω_0^2 to be greater than K^2 (i.e. weak damping), the general solution of Eq. (6.2-2) will be of the form

$$x(t)=Ae^{-Kt}\cos[\sqrt{\omega_0^2-K^2}\,t-\theta]+a\cos(\omega t-\varphi). \tag{6.2-10}$$

The first term on the RHS represents the transient solution (corresponding to the natural vibrations of the system) which eventually die off. The second term represents the steady state solution which corresponds to the forced vibrations imposed by the external force. Notice that the frequency of the forced vibrations is the same as that of the external force.

6.3 Resonance

The amplitude of the forced vibration

$$a=\frac{G}{[(\omega_0^2-\omega^2)^2+4K^2\omega^2]^{1/2}} \tag{6.3-1}$$

depends on the frequency of the driving force and is a maximum when

$$(\omega_0^2-\omega^2)^2+4K^2\omega^2$$

is a minimum; i.e. when

$$\frac{d}{d\omega}[(\omega_0^2-\omega^2)^2+4K^2\omega^2]=0$$

or

$$2(\omega_0^2-\omega^2)(-2\omega)+8K^2\omega=0$$

or

$$\omega=\omega_0\left[1-\frac{2K^2}{\omega_0^2}\right]^{1/2} \tag{6.3-2}$$

Thus the amplitude is maximum* when ω is given by Eq. (6.3-2); this is known as amplitude resonance. When the damping is extremely small, the resonance occurs at a frequency very close to the natural frequency of the system. The variation of the amplitude with ω is shown in Fig. 6.3. Notice that as the damping decreases, the maximum becomes very sharp and the amplitude falls off rapidly as we go away from the resonance. The maximum value of a is given by

$$a_{\max}=\frac{G}{\left[(2K^2)^2+4K^2\omega_0^2\left(1-\frac{2K^2}{\omega_0^2}\right)\right]^{1/2}}$$

$$=\frac{G}{2K[\omega_0^2-K^2]^{1/2}}=\frac{G}{2K[\omega^2+K^2]^{1/2}} \qquad (6.3\text{-}3)$$

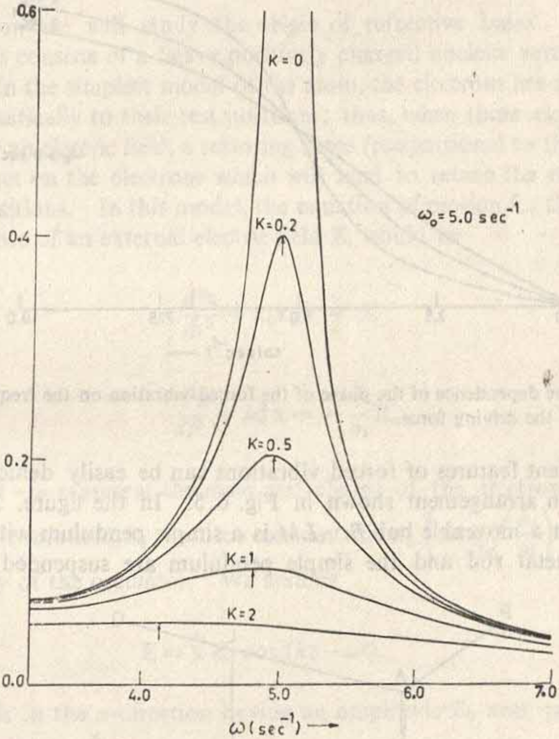


Fig. 6.3 The variation of amplitude with the frequency of the external driving force for various values of K. The calculations correspond to $\omega_0 = 5$ sec^{-1} and the values of K are in sec^{-1}. Notice that with increase in damping the resonance occurs at a smaller value of ω. The arrows denote the positions of the maxima.

*There is no resonance condition when $K^2 \geqslant \frac{1}{2}\omega_0^2$.

Thus, with increase in damping, the maximum occurs at lower values of ω and the resonance becomes less sharper.

In order to discuss the phase of the forced vibrations we refer to Eq. (6.2-9) from where we find that for small damping the phase angle is small unless it is near resonance. For $\omega = \omega_0$, $\tan\varphi = \infty$ and φ is $\pi/2$; i.e. the phase of the forced vibration is $\pi/2$ *ahead* of the phase of the driving force. As the frequency of the driving force is increased beyond ω_0, the phase also increases and approaches π (Fig. 6.4).

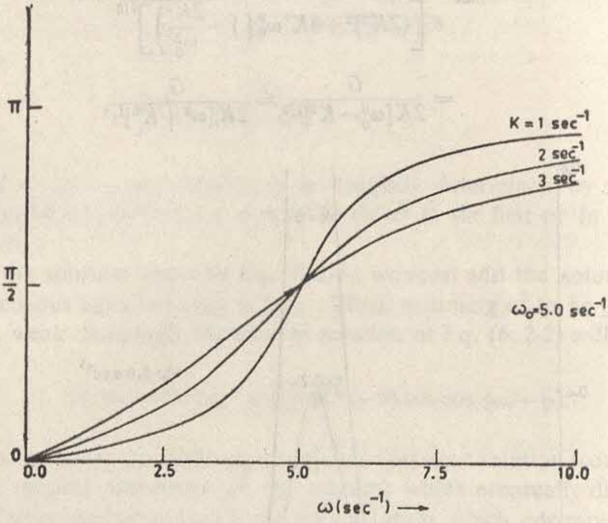


Fig. 6.4 The dependence of the phase of the forced vibration on the frequency of the driving force.

All the salient features of forced vibrations can be easily demonstrated by means of an arrangement shown in Fig. 6.5. In the figure, *AC* is a metal rod with a moveable bob *B*. *LM* is a simple pendulum with a bob at *M*. The metal rod and the simple pendulum are suspended from a

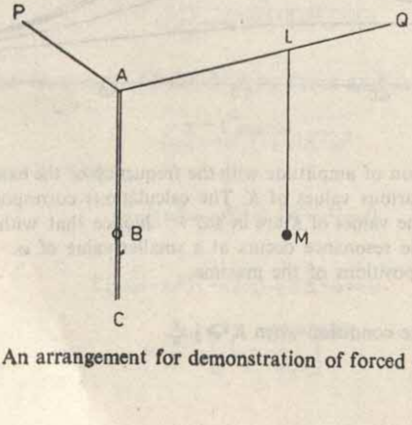


Fig. 6.5 An arrangement for demonstration of forced vibrations.

string *PQ* as shown in Fig. 6.5. With *B* at the bottom, when the rod *AC* is set in motion, the pendulum *LM* also vibrates. As the bob *B* is moved upwards the time period decreases and the frequency of the rod becomes closer to the natural frequency of the simple pendulum and eventually the resonance condition is satisfied. At resonance the amplitude of vibration of the simple pendulum is maximum and the phase difference between the vibrations is nearly $\pi/2$; i.e. when the metal rod is at its lowest position and moving towards right, the simple pendulum is at the extreme left position.

If the bob *B* is further moved upward, the frequency increases and the amplitude of the forced vibrations decreases.

6.4 Origin of Refractive Index

In this section we will study the origin of refractive index. We know that an atom consists of a heavy positively charged nucleus surrounded by electrons. In the simplest model of the atom, the electrons are assumed to be bound elastically to their rest positions; thus, when these electrons are displaced by an electric field, a restoring force (proportional to the displacement) will act on the electrons which will tend to return the electrons to their rest positions. In this model, the equation of motion for the electron, in the presence of an external electric field **E**, would be

$$m\frac{d^2\mathbf{x}}{dt^2}+k_0\mathbf{x}=-e\mathbf{E} \tag{6.4-1}$$

or

$$\frac{d^2\mathbf{x}}{dt^2}+\omega_0^2\mathbf{x}=-\frac{e}{m}\mathbf{E}, \tag{6.4-2}$$

where m and $-e$ represent the mass and charge of the electron ($e>0$), **E** the external electric field, k_0 the force constant and $\omega_0\left(=\sqrt{\frac{k_0}{m}}\right)$ represents the frequency of the oscillator. We assume

$$\mathbf{E}=\hat{\mathbf{x}}\,E_0\cos(kz-\omega t), \tag{6.4-3}$$

i.e. the field is in the x-direction having an amplitude E_0 and propagating in the $+z$-direction; $\hat{\mathbf{x}}$ represents the unit vector in the x-direction. Thus

$$\frac{d^2x}{dt^2}+\omega_0^2x=-\frac{eE_0}{m}\cos(kz-\omega t), \tag{6.4-4}$$

where we have replaced the vectors by the corresponding scalar quantities because the displacement and the electric field are in the same direction. Except for the damping term, Eq. (6.4-4) is similar to Eq. (6.2-1) and, therefore, the solution corresponding to the forced vibrations will be given by*

$$x=-\frac{eE_0}{m(\omega_0^2-\omega^2)}\cos(kz-\omega t) \tag{6.4-5}$$

In the simplest model of the atom, the centre of the negative charge (due to the electrons) is assumed to be at the nucleus. In the presence of an electric field, the centre of the negative charge gets displaced from the nucleus which results in a finite value of the dipole moment of the atom. In particular, if we have a positive charge $+e$ at the origin and a negative charge $-e$ at a distance x, then the dipole moment would be $-ex$; thus, if there are N dispersion electrons** per unit volume then the polarization (i.e. dipole moment per unit volume) would be given by

$$\begin{aligned}\mathbf{P}=-Ne\,\mathbf{x}&=\frac{Ne^2}{m(\omega_0^2-\omega^2)}\mathbf{E}\\&=\chi\mathbf{E}\end{aligned} \tag{6.4-6}$$

where

$$\chi=\frac{N\epsilon}{m(\omega_0^2-\omega^2)} \tag{6.4-7}$$

is known as the electric susceptibility of the material. The dielectric permittivity is therefore given by (see Chapter 19)

$$\epsilon=\epsilon_0+\chi, \tag{6.4-8}$$

or

$$\frac{\epsilon}{\epsilon_0}=1+\frac{Ne^2}{m\,\epsilon_0\,\omega_0^2}\left[1-\frac{\omega^2}{\omega_0^2}\right]^{-1} \tag{6.4-9}$$

Now, ϵ/ϵ_0 is the dielectric constant κ, which is equal to the square of the refractive index (see Chapter 19). Thus

$$n^2=1+\frac{Ne^2}{m\,\epsilon_0\,\omega_0^2}\left[1-\frac{\omega^2}{\omega_0^2}\right]^{-1} \tag{6.4-10}$$

showing that the refractive index depends on the frequency; this is known as dispersion. Assuming that the characteristic frequency ω_0 lies in the far ultraviolet (see Eq. 6.4-16)***, the quantity $\left[1-\frac{\omega^2}{\omega_0^2}\right]^{-1}$ is positive in the entire visible region. Further, as ω increases n^2 also increases; i.e. the refractive index increases with frequency, this is known as normal dispersion. If we further assume $\frac{\omega}{\omega_0} \ll 1$, then

---

*Notice that in the absence of damping (i.e. when $\Gamma=0$), $\varphi=0$; see Eq. (6.2-9).

**The number of "dispersion-electrons" in a molecule of an ideal gas is the valence number of the molecules. This number is 2 for H_2, 6 for N_2, etc.

***This also follows from the fact that according to classical electrodynamics, an oscillating dipole (vibrating with frequency ω_0) will radiate electromagnetic waves with frequency ω_0; and as an example if we consider hydrogen, then $\hbar\omega_0\approx 13.6\ eV$ from which one obtains $\omega_0\approx 2\times10^{16}\ \text{sec}^{-1}$. This frequency corresponds to the far ultraviolet.

$$\left[1-\frac{\omega^2}{\omega_0^2}\right]^{-1} \approx 1+\frac{\omega^2}{\omega_0^2},$$

and

$$n^2 \simeq 1+\frac{Ne^2}{m\epsilon_0\omega_0^2}\left[1+\frac{\omega^2}{\omega_0^2}\right]$$

$$\simeq 1+\frac{Ne^2}{m\epsilon_0\omega_0^2}+\frac{4\pi^2c^2Ne^2}{m\epsilon_0\omega_0^4}\cdot\frac{1}{\lambda_0^2}, \qquad (6.4\text{-}11)$$

where $\lambda_0 = 2\pi c/\nu$ is the free space wavelength. Eq. (6.4-11) can be written in the form

$$n^2 = A + B/\lambda_0^2, \qquad (6.4\text{-}12)$$

which is the well known Cauchy relation. For hydrogen,* the experimental variation of n^2 with λ_0 is given by the following relation

$$n^2 = 1 + 2.721\times10^{-4} + \frac{2.11\times10^{-18}}{\lambda_0^2} \qquad (6.4\text{-}13)$$

where the wavelength is measured in meters. Thus

$$\frac{Ne^2}{m\epsilon_0\omega_0^2} = 2.721\times10^{-4} \qquad (6.4\text{-}14)$$

and

$$\frac{4\pi^2c^2Ne^2}{m\epsilon_0\omega_0^4} = 2.11\times10^{-18}m^2 \qquad (6.4\text{-}15)$$

If we divide Eq. (6.4-15) by Eq. (6.4-14), we would get

$$\frac{4\pi^2c^2}{\omega_0^2} = \frac{2.11\times10^{-18}}{2.721\times10^{-4}}$$

or,

$$\nu_0 = \frac{\omega_0}{2\pi} = \left[\frac{2.721\times10^{-4}}{2.11\times10^{-18}}\right]^{1/2} c$$

$$\approx 3\times10^{15}\,\text{sec}^{-1} \qquad (6.4\text{-}16)$$

Thus

$$\lambda_0 \approx 10^{-7}\,\text{m} = 10^{-5}\,\text{cm}$$

which is indeed in the ultraviolet region. One can eliminate ω_0 from Eqs. (6.4-14) and (6.4-15) to obtain

---

*Data corresponds to 0°C and 76 cm of Hg[see, for example, Sommerfeld (1964)].

$$\frac{\left(\dfrac{Ne^2}{m\epsilon_0}\right)^2}{4\pi^2c^2Ne^2/m\,\epsilon_0}=\frac{2.721\times2.721\times10^{-8}}{2.11\times10^{-18}}$$

or

$$\frac{Ne^2}{4\pi^2c^2\epsilon_0 m}\approx3\times10^{10}\,\mathrm{m}^{-2} \tag{6.4-17}$$

Now at *NTP*, 22400 cc of H_2 contains 6×10^{23} molecules; thus

$$\begin{aligned}N&=2\times\frac{6\times10^{23}}{22400\times10^{-6}}\,m^{-3}\\&\approx5\times10^{25}m^{-3}\end{aligned}$$

where the factor 2 arises from the fact that a hydrogen molecule consists of two electrons.* Hence

$$\begin{aligned}\frac{Ne^2}{4\pi^2c^2\epsilon_0 m}&\simeq\frac{5\times10^{25}\times(1.6\times10^{-19})^2}{4\times10\times9\times10^{16}\times8.85\times10^{-12}\times9.1\times10^{-31}}\\&\simeq4\times10^{10}m^{-2}\end{aligned} \tag{6.4-18}$$

which qualitatively agrees with Eq. (6.4-17).

It is of interest to mention that for a gas of free electrons (as we have in the upper atmosphere) there is no restoring force and we must set $\omega_0=0$. Thus the expression for the refractive index becomes

$$n^2=1-\frac{Ne^2}{m\epsilon_0\omega^2} \tag{6.4-19}$$

which shows that the refractive index is less than unity.** Indeed the above equation gives the correct dependence of the refractive index of the stratosphere for radiowaves.*** In Eq. (6.4-19), N represents the density of free electrons.

Returning to Eq. (6.4-10), we note that as $\omega\rightarrow\omega_0$, the refractive index tends to ∞. This is due to the fact that we have neglected the presence of damping forces in our treatment. If we do take into account the damping forces, Eq. (6.4-4) would modify to (see Eq. 6.2-1)

$$m\frac{d^2x}{dt^2}+\Gamma\frac{dx}{dt}+k_0x=eE_0\cos(kz-\omega t) \tag{6.4-20}$$

---

*See footnote (**) on p. 112.

**This does not, however, imply that one can send signals faster than the speed of light in free space (see Chapter 7).

***In plasma physics books, Eq. (6.4-19) is written in the form

$$n^2=1-\left(\frac{\omega_p}{\omega}\right)^2$$

where $\omega_p=\left(\dfrac{Ne^2}{m\epsilon_0}\right)^{1/2}$ is known as the plasma frequency. Notice that for $\omega<\omega_p$, the refractive index is pure imaginary which gives rise to attenuation.

In order to derive an expression for the refractive index, it is more convenient to rewrite Eq. (6.4-20) in the form

$$\frac{d^2x}{dt^2}+2K\frac{dx}{dt}+\omega_0^2x=\frac{eE_0}{m}\,e^{i(kz-\omega t)} \tag{6.4-21}$$

where the solution of Eq. (6.4-20) will be the real part of the solution of Eq. (6.4-21). The solution of the homogeneous equation will give the transient behaviour which will die out as $t\rightarrow\infty$ (see Sec. 6.2); the steady state solution will correspond to frequency ω. Thus, if we substitute a solution of the type

$$x(t)=Ae^{i(kz-\omega t)} \tag{6.4-22}$$

in Eq. (6.4-21), we would obtain

$$(-\omega^2-2iK\omega+\omega_0^2)\,A=\frac{eE_0}{m}$$

or*

$$A=\frac{eE_0}{m\,[\omega_0^2-\omega^2-2iK\omega]} \tag{6.4-23}$$

Thus, we get

$$\mathbf{P}=\frac{Ne^2}{m\,[\omega_0^2-\omega^2-2iK\omega]}\,\mathbf{E} \tag{6.4-24}$$

The electric susceptibility would therefore be given by

$$\chi=\frac{Ne^2}{m\,(\omega_0^2-\omega^2-2iK\omega)}$$

Thus

$$n^2=\frac{\epsilon}{\epsilon_0}=1+\frac{\chi}{\epsilon_0}$$

$$=1+\frac{Ne^2}{m\epsilon_0\,[\omega_0^2-\omega^2-2iK\omega]} \tag{6.4-25}$$

Notice that the refractive index is complex, which implies absorption of the propagating electromagnetic wave. Indeed, if we write

$$n=\eta+i\kappa \tag{6.4-26}$$

where η and κ are real numbers, then the wave number k, which equals $n\omega/c$, would be given by

$$k=(\eta+i\kappa)\,\frac{\omega}{c}\cdot \tag{6.4-27}$$

---

*Notice that A is complex; however, if we substitute the expression for A from Eq. (6.4-23) in Eq. (6.4-22) and take the real part we would get the same expression for $x(t)$ as we had obtained in Sec. 6.2.

If we consider a plane electromagnetic wave propagating in the $+z$ direction, then its z and t dependence would be of the form $\exp[i(kz-\omega t)]$; consequently

$$\begin{aligned}\mathbf{E} &= \mathbf{E}_0 e^{i(kz-\omega t)}\\ &= \mathbf{E}_0 \exp\left[i\left\{(\eta+i\kappa)\frac{\omega}{c}z-\omega t\right\}\right]\\ &= \mathbf{E}_0 \exp\left[-i\omega\left(t-\frac{\eta z}{c}\right)-\frac{\kappa\omega}{c}z\right]\end{aligned} \tag{6.4-28}$$

which shows an exponential attenuation of the amplitude. This should not be unexpected because damping causes a loss of energy.

In order to obtain expressions for η and κ, we substitute the expression for n from Eq. (6.4-26) in Eq. (6.4-25) to obtain

$$(\eta+i\kappa)^2 = 1+\frac{Ne^2}{m\epsilon_0(\omega_0^2-\omega^2-2iK\omega)}\frac{(\omega_0^2-\omega^2+2iK\omega)}{(\omega_0^2-\omega^2+2iK\omega)}$$

or

$$\eta^2-\kappa^2 = 1+\frac{Ne^2(\omega_0^2-\omega^2)}{m\epsilon_0[(\omega_0^2-\omega^2)^2+4K^2\omega^2]} \tag{6.4-29}$$

and

$$2\eta\kappa = \frac{Ne^2}{m\epsilon_0}\frac{2K\omega}{[(\omega_0^2-\omega^2)^2+4K^2\omega^2]} \tag{6.4-30}$$

The above equations can be rewritten in the form

$$\eta^2-\kappa^2 = 1-\frac{\alpha\Omega}{[\Omega^2+\beta^2(\Omega+1)]} \tag{6.4-31}$$

and

$$2\eta\kappa = \frac{\alpha\beta\sqrt{1+\Omega}}{[\Omega^2+\beta^2(\Omega+1)]} \tag{6.4-32}$$

where we have introduced the following dimensionless parameters:

$$\alpha = \frac{Ne^2}{m\epsilon_0\omega_0^2};\ \Omega = \frac{\omega^2-\omega_0^2}{\omega_0^2}$$

and

$$\beta = \frac{2K}{\omega_0^2}$$

The qualitative variations of $\eta^2-\kappa^2$ and $2\eta\kappa$ with Ω are shown in Fig. 6.6. It can easily be shown that at $\Omega=-\beta$ and at $\Omega=+\beta$, the function $(\eta^2-\kappa^2)$ attains its maximum and minimum values respectively.

It should be pointed out that, in general, an atom can execute oscillations corresponding to different resonant frequencies and we have to take into account the various contributions. If $\omega_0, \omega_1, \ldots$ represent the resonant frequencies and if f_j represents the fractional number of

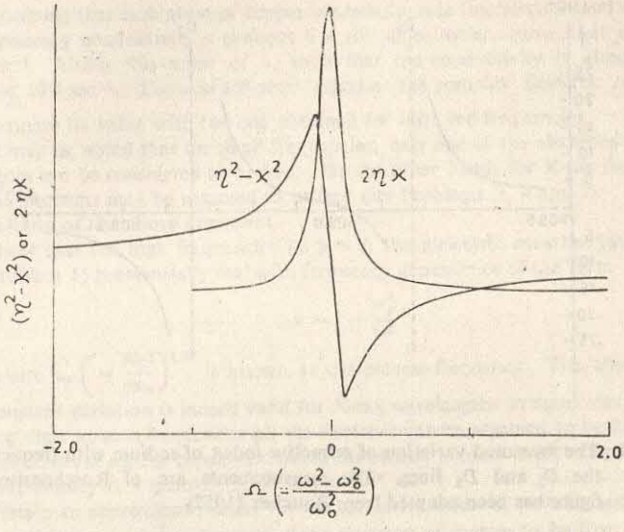


Fig. 6.6 Qualitative variation of $(\eta^2-\kappa^2)$ and $2\eta\kappa$ with $\boldsymbol{\Omega}$.

electrons per unit volume whose resonant frequency is ω_j, then Eq. (6.4-25) would get modified to the following expression:*

$$n^2 = 1 + \frac{Ne^2}{m\epsilon_0}\sum_j \frac{f_j}{[\omega_j^2-\omega^2-2iK_j\omega]} \tag{6.4-33}$$

where K_j represents the damping constant corresponding to the resonant frequency ω_j. Indeed, Eq. (6.4-33) describes correctly the variation of refractive index for most gases. Fig. 6.7 shows the dependence of the refractive index of sodium vapour around $\lambda_0 = 5890$ Å. Since D_1 and D_2 lines occur at 5890 Å and 5896 Å, one should expect resonant oscillations around these frequencies. This is indeed borne out by the data shown in Fig. 6.7. The variation of the refractive index can be accurately fitted with the formula

$$n^2 = 1 + \frac{A^2}{\nu^2-\nu_1^2} + \frac{B}{\nu^2-\nu_2^2} \tag{6.4-34}$$

where we have neglected the presence of damping forces which is justified except when one is very close to the resonance.

It may be worthwhile mentioning that in a liquid the molecules are very close to one another and the dipoles interact between themselves. If we take this interaction into account, we would get**

---

*Quantum mechanics also gives a similar result (see, for example, R. Loudon *The Quantum Theory of Light*, Clarendon Press, Oxford, 1973).

**See, for example, Bottcher (1952). Notice that when n is very close to unity (i.e, for a dilute fluid), Eq. (6.4-35) reduces to Eq. (6.4-33).

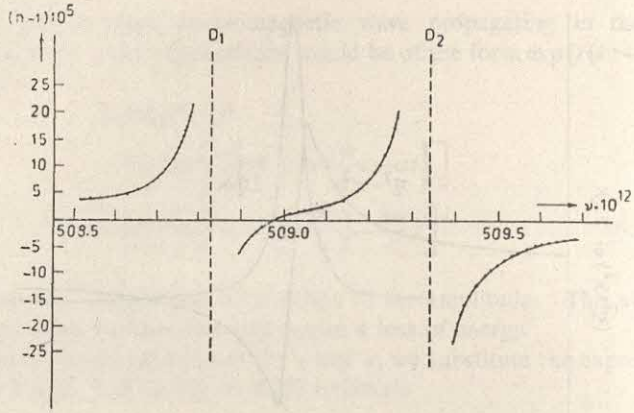


Fig. 6.7 The measured variation of refractive index of sodium with frequency around the D_1 and D_2 lines. The measurements are of Roschdestwensky; the figure has been adapted from Bottcher (1952).

$$\frac{n^2-1}{n^2+2}=\frac{Ne^2}{3m\epsilon_0}\sum_j\frac{f_l}{\omega_j^2-\omega^2} \tag{6.4-35}$$

where we have neglected the presence of damping. For liquids, whose molecules do not have a permanent dipole moment (e.g. H_2, O_2, etc.) Eq. (6.4-35) gives a fairly accurate description. However, liquids whose molecules possess permanent dipole moments (e.g. H_2O) one has to carry out a different analysis.

PROBLEMS

1. (a) In a metal, the electrons can be assumed to be essentially free. Show that the drift velocity of the electron satisfies the following equation

$$m\frac{d\mathbf{v}}{dt}+m\nu\mathbf{v}=\mathbf{F}=-e\mathbf{E}\,e^{-i\omega t}$$

where ν represents the collision frequency. Show that the above equation implies the randomization of the momentum after each collision.

(b) Calculate the steady state current density ($\mathbf{J}=-Ne\mathbf{v}$) and show that the conductivity is given by

$$\sigma(\omega)=\frac{Ne^2}{m}\frac{1}{\nu-i\omega}$$

(c) If $\mathbf{r}$ represents the displacement of the electron, show that

$$\mathbf{P}=-Ne\mathbf{r}=-\frac{Ne^2}{m(\omega^2+i\omega\nu)}\mathbf{E}$$

which represents the polarization. Using the above equation show that

$$\kappa(\omega)=1-\frac{Ne^2}{m\epsilon_0(\omega^2+i\omega\nu)}$$

which represents the dielectric constant variation for a free electron gas.

2. Assuming that each atom of copper contributes one free electron and that the low frequency conductivity σ is about 6×10^7 mhos/meter, show that $\nu \simeq 3 \times 10^{13}$ sec^{-1}. Using this value of ν, show that the conductivity is almost real for $\omega \lesssim 10^{11}$ sec^{-1}. For $\omega = 10^8$ sec^{-1} calculate the complex dielectric constant and compare its value with the one obtained for infra-red frequencies.
   It may be noted that for small frequencies, only one of the electrons of a copper atom can be considered to be free. On the other hand, for X-ray frequencies all the electrons may be assumed to be free (see Problems 3, 4 and 5). Discuss the validity of the above argument.
3. Show that for high frequencies ($\omega >> \nu$) the dielectric constant (as derived in Problem 1) is essentially real with frequency dependence of the form

$$\kappa = 1 - \frac{\omega_p^2}{\omega^2}$$

   where $\omega_p \left(= \frac{Ne^2}{m\epsilon_0} \right)^{1/2}$ is known as the plasma frequency. The above dielectric constant variation is indeed valid for X-ray wavelengths in many metals. Assuming that at such frequencies all the electrons can be assumed to be free, calculate the constant ω_p for copper for which the atomic number is 29, mass number is 63 and density is 9 g/cm^3. (Ans: $\sim 3 \times 10^{16}$ sec^{-1})
4. Obtain an approximate value for the refractive index of metallic sodium corresponding to $\lambda = 1$ Å. Assume all the electrons of sodium to be free.
5. Problem 3 tells us that for $\omega > \omega_p$, the refractive index is real. Indeed in 1933, Wood discovered that alkali metals are transparent to ultraviolet light. Assuming that the refractive index is primarily due to the free electrons (and that there is one free electron per atom) show that the values of the wave length $\lambda_p \left(= \frac{2\pi c}{\omega_p} \right)$ for Li, Na, K, Rb are 1551 Å, 2093 Å, 2884 Å and 3214 Å respectively;* the atomic weights of Li, Na, K, Rb, are 6.94, 22.99, 39.10 and 85.48 respectively; the densities are 0.534, 0.9712, 0.870 and 1.532 g/cm^3; $m = 9.109 \times 10^{-31}$ kg, $e = 1.602 \times 10^{-19}$ C and $\epsilon_0 = 8.854 \times 10^{-12}$ C/N-m^2.
6. In an ionic crystal (like NaCl, CaF$_2$, etc.) one has to take into account infra-red resonance oscillations of the ions. Show that Eq. (6.4-10) modifies to

$$n^2 = 1 + \frac{Ne^2}{m\epsilon_0 (\omega_1^2 - \omega^2)} + \frac{p\,Ne^2}{M\epsilon_0 (\omega_2^2 - \omega^2)}$$

   where M represents the reduced mass of the two ions and p represents the valence of the ion ($p = 1$ for Na$^+$, Cl$^-$, $p = 2$ for Ca^{++}, F$_2^{--}$). Show that the above equation can be written in the form

$$n^2 = n_\infty^2 + \frac{A_1}{\lambda^2 - \lambda_1^2} + \frac{A_2}{\lambda^2 - \lambda_2^2}$$

   where

$$\lambda_1 = \frac{2\pi c}{\omega_1}, \quad \lambda_2 = \frac{2\pi c}{\omega_2}$$

$$A_1 = \frac{Ne^2}{4\pi^2 c^2 \epsilon_0 m} \lambda_1^4, \quad A_2 = \frac{p\,Ne^2}{4\pi^2 c^2 \epsilon_0 M} \lambda_2^4$$

7. The refractive index variation for CaF$_2$ (in the visible region of the spectrum) can be written in the form**

$$n^2 = 6.09 + \frac{6.12 \times 10^{-15}}{\lambda^2 - 8.88 \times 10^{-15}} + \frac{5.10 \times 10^{-9}}{\lambda^2 - 1.26 \times 10^{-9}}$$

   where λ is in meters.

---

*The corresponding experimental values are 1550 Å, 2100 Å, 3150 Å and 3400 Å.

**Quoted from Sommerfeld (1964); measurements are of Paschen,

(a) Plot the variation of n^2 with λ in the visible region.
(b) From the values of A_1 and A_2 show that $m/M \approx 2.07 \times 10^{-5}$ and compare this with the exact value.
(c) Show that the value of n_∞ obtained by using the constants A_1, A_2, λ_1 and λ_2 agrees reasonably well with the experimental value.

8. (a) The refractive index of a plasma (neglecting the collisions) is approximately given by (see Sec. 6·4)

$$n^2 = 1 - \frac{\omega_p^2}{\omega^2}$$

where

$$\omega_p = \left(\frac{Ne^2}{m\varepsilon_0}\right)^{1/2} \simeq 56.414\, N^{1/2}\, \text{sec}^{-1}$$

is known as the plasma frequency. In the ionosphere the maximum value of N_0 is $\simeq 10^{10} - 10^{12}$ electrons/m³. Calculate the plasma frequency. Notice that at high frequencies $n^2 \simeq 1$; thus high frequency waves (like the one used in TV) are not reflected by the ionosphere. On the other hand, for low frequencies, the refractive index is imaginary (like in a conductor—see Sec. 20.3) and the beam gets reflected. This fact is used in long distance radio communications.

(b) Assume that for $x \simeq 200$ km, $N = 10^{12}$ electrons/m³ and that the electron density increases to 2×10^{12} electrons/m³ at $x \simeq 300$ km. For $x \gtrsim 300$ km, the electron density decreases. Assuming a parabolic variation of N, plot the corresponding refractive index variation.

9. In the low frequency region, the refractive index of water decreases from about 9.0 at room temperature to about 7.5 near the boiling point of water. On the other hand, in the optical region, the refractive index is about 1.33 and changes by about 1% in the above temperature range. Explain this behaviour qualitatively. (Hint: Water molecules have a permanent dipole moment which are randomly oriented in the absence of an electric field).

REFERENCES AND SUGGESTED READING

See at the end of Chapter 7.

7 Fourier Series and Applications

Again! Again! Again!

SHAKESPEARE

7.1 Fourier Analysis

According to Fourier's theorem any periodic vibration can be expressed as a sum of sine and cosine functions whose frequencies increase in the ratio of natural numbers. Thus, a periodic function with period T, i.e.

$$f(t+nT)=f(t);\ n=0,\pm 1,\pm 2,\ldots; \tag{7.1-1}$$

can be expanded* in the form:

$$\begin{aligned} f(t) &= \tfrac{1}{2}a_0+\sum_{n=1}^{\infty} a_n\cos\left(\frac{2n\pi}{T}\,t\right)+\sum_{n=1}^{\infty} b_n\sin\left(\frac{2n\pi}{T}\,t\right) \\ &= \tfrac{1}{2}a_0+\sum_{n=1}^{\infty} a_n\cos(n\,\omega t)+\sum_{n=1}^{\infty} b_n\sin(n\,\omega t) \end{aligned} \tag{7.1-2}$$

where
$$\omega=\frac{2\pi}{T} \tag{7.1-3}$$

---

*Actually, for the expansion to be possible, the function $f(t)$ must satisfy certain conditions. The conditions are that the function $f(t)$ in one period (i.e. in the interval $t_0<t<t_0+T$) must be (i) single valued, (ii) piecewise continuous (i.e. it can have at most a finite number of finite discontinuities) and (iii) can have only a finite number of maxima and minima. These conditions are known as Dirichlet's conditions and are almost always satisfied in all problems that one encounters in physics.

represents the fundamental frequency.

The coefficients a_n and b_n can easily be determined by using the following properties of the trigonometric functions:

$$\int_{t_0}^{t_0+T} \cos n\omega t \ \cos m\omega t \, dt = \begin{cases} 0 & \text{if } m \neq n \\ T/2 & \text{if } m = n \end{cases} \tag{7.1-4}$$

$$\int_{t_0}^{t_0+T} \sin n\omega t \ \sin m\omega t \, dt = \begin{cases} 0 & \text{if } m \neq n \\ T/2 & \text{if } m = n \end{cases} \tag{7.1-5}$$

$$\int_{t_0}^{t_0+T} \sin n\omega t \ \cos m\omega t \, dt = 0. \tag{7.1-6}$$

The above equations can easily be derived. For example, for $m = n$

$$\begin{aligned} \int^{t_0+T} \cos n\omega t \ \cos m\omega t \, dt &= \int_{t_0}^{t_0+T} \cos^2 n\omega t \, dt \\ &= \frac{1}{2} \int_{t_0}^{t_0+T} [1 + \cos 2n\,\omega t] \, dt \\ &= \frac{1}{2} \Big[T + \frac{1}{2n\omega} \{\sin (2n\omega t_0 + 4n\pi) \\ &\qquad - \sin (2n\omega t_0)\} \Big] \\ &= \frac{T}{2}. \end{aligned}$$

Similarly, for $m \neq n$

$$\begin{aligned} &\int_{t_0}^{t_0+T} \cos n\omega t \ \cos m\omega t \, dt \\ &= \frac{1}{2} \int_{t_0}^{t_0+T} [\cos (n-m)\,\omega t + \cos (n+m)\,\omega t] \, dt \\ &= \frac{1}{2} \left[\frac{1}{(n-m)\,\omega} \sin (n-m)\,\omega t + \frac{1}{(n+m)\,\omega} \sin (n+m)\,\omega t \right]_{t_0}^{t_0+T} \\ &= 0. \end{aligned}$$

In order to determine the coefficients a_n and b_n we first multiply Eq. (7.1-2) by dt and integrate from t_0 to $t_0 + T$:

$$\int_{t_0}^{t_0+T} f(t)\,dt = \tfrac{1}{2}a_0 \int_{t_0}^{t_0+T} dt + \sum_{n=1}^{\infty} a_n \int_{t_0}^{t_0+T} \cos n\omega t\,dt + \sum_{n=1}^{\infty} b_n \int_{t_0}^{t_0+T} \sin n\omega t\,dt$$

$$= \frac{T}{2}a_0$$

where we have used Eqs. (7.1-4) and (7.1-6) for $m = 0$. Thus

$$a_0 = \frac{2}{T}\int_{t_0}^{t_0+T} f(t)\,dt \tag{7.1-7}$$

Next, if we multiply Eq. (7.1-2) by $\cos(m\omega t)\,dt$ and integrate from t_0 to $t_0 + T$, we would obtain

$$\int_{t_0}^{t_0+T} f(t)\cos(m\omega t)\,dt = \tfrac{1}{2}a_0 \int_{t_0}^{t_0+T} \cos(m\omega t)\,dt + \sum_{n=1}^{\infty} a_n \int_{t_0}^{t_0+T} \cos(m\omega t)\cos(n\omega t)\,dt + \sum b_n \int_{t_0}^{t_0+T} \cos(m\omega t)\sin(n\omega t)\,dt$$

$$= \frac{T}{2}a_m$$

where we have used Eqs. (7.1-4) and (7.1-6). We may combine the above equation with Eq. (7.1-7) to write

$$a_n = \frac{2}{T}\int_{t_0}^{t_0+T} f(t)\cos n\omega t\,dt;\quad n = 0, 1, 2, 3, \ldots. \tag{7.1-8}$$

Similarly

$$b_n = \frac{2}{T}\int_{t_0}^{t_0+T} f(t)\sin n\omega t\,dt;\quad n = 1, 2, 3, \ldots. \tag{7.1-9}$$

It should be pointed out that the value of t_0 is quite arbitrary. In some problems it is convenient to choose

$$t_0 = -T/2$$

then

$$a_n = \frac{2}{T}\int_{-T/2}^{+T/2} f(t)\cos n\omega t\,dt;\ \ n=0,1,2\ldots$$

and

$$b_n = \frac{2}{T}\int_{-T/2}^{+T/2} f(t)\sin n\omega t\,dt;\ \ n=0,1,2\ldots$$

Such a choice is particularly convenient when the function is even (i.e. $f(t)=f(-t)$) or odd (i.e. $f(t)=-f(-t)$). In the former case $b_n=0$ whereas in the latter case $a_n=0$. Often it is convenient to choose $t_0=0$.

Example 7.1 Consider a periodic function of the form

$$\left.\begin{aligned} f(t) &= t \qquad \text{for} \quad -\tau < t < +\tau \\ \text{and} \qquad f(t+2n\tau) &= f(t) \end{aligned}\right\} \qquad (7.1\text{-}10)$$

(see Fig. 7.1). Such a function is referred to as a saw tooth function. Expand the above function in a Fourier series.

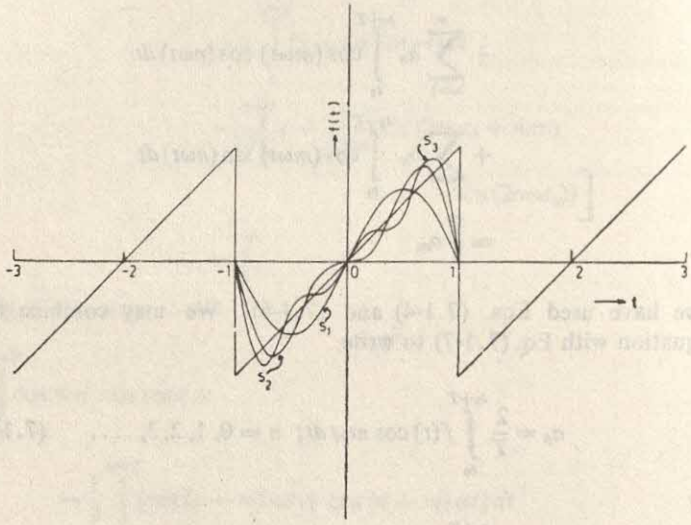


Fig. 7·1 The saw tooth function. S_1, S_2 and S_3 represent the partial sums corresponding to the saw tooth function.

Solution. Since $f(t)$ is an odd function of t, $a_n=0$ and

$$b_n = \frac{2}{T}\int_{-\pi}^{+\pi} f(t)\sin(n\omega t)\,dt$$

$$= \frac{1}{\tau}\, 2 \int_0^{\pi} t \sin n\omega t \, dt$$

Notice that the periodicity is 2τ and, therefore, $\omega = \frac{\pi}{\tau}$. Carrying out the integration we obtain

$$\begin{aligned} b_n &= \frac{2}{\tau}\left[-\frac{t}{n\omega}\cos n\omega t + \frac{1}{n\omega}\left\{\frac{1}{n\omega}\sin n\omega t\right\}\right]_0^{\tau} \\ &= -\frac{2\tau}{n\pi}\cos n\pi = (-1)^{n+1}\frac{2\tau}{n\pi} \end{aligned} \qquad (7.1\text{-}11)$$

Thus

$$\begin{aligned} f(t) &= \frac{2\tau}{\pi}\sum_{n=1,2\ldots}\frac{(-1)^{n+1}}{n}\sin n\omega t \\ &= \frac{2\tau}{\pi}\left[\sin \omega t - \frac{1}{2}\sin 2\omega t + \frac{1}{3}\sin 3\omega t - \ldots\right] \end{aligned} \qquad (7.1\text{-}12)$$

In Fig. 7.1 we have also plotted the partial sums which are given by

$$S_1 = \frac{2\tau}{\pi}\sin \omega t; \quad S_2 = \frac{2\tau}{\pi}\left[\sin \omega t - \frac{1}{2}\sin 2\omega t\right]$$

$$S_3 = \frac{2\tau}{\pi}\left[\sin \omega t - \frac{1}{2}\sin 2\omega t + \frac{1}{3}\sin 3\omega t\right]$$

It can be seen from the figure that as n increases the sum S_n approaches the function $f(t)$.

Example 7.2 Fourier expand the function defined by the following equations

$$\begin{aligned} f(t) &= -A \quad \text{for} \quad -\frac{T}{2} < t < 0 \\ &= +A \quad \text{for} \quad 0 < t < +\frac{T}{2} \end{aligned} \qquad (7.1\text{-}13)$$

and

$$f(t+T) = f(t)$$

Solution. The function is plotted in Fig. 7.2. Once again the function is an odd function, consequently $a_n = 0$ and

$$\begin{aligned} b_n &= \frac{2}{T}2\int_0^{T/2} A \sin(n\omega t)\, dt = \frac{4A}{T}\frac{1}{n\omega}\Big[-\cos n\omega t\Big]_0^{T/2} \\ &= \frac{2A}{n\pi}\Big[1 - \cos n\pi\Big] \\ &= \frac{2A}{n\pi}\Big[1 - (-1)^n\Big] \end{aligned}$$

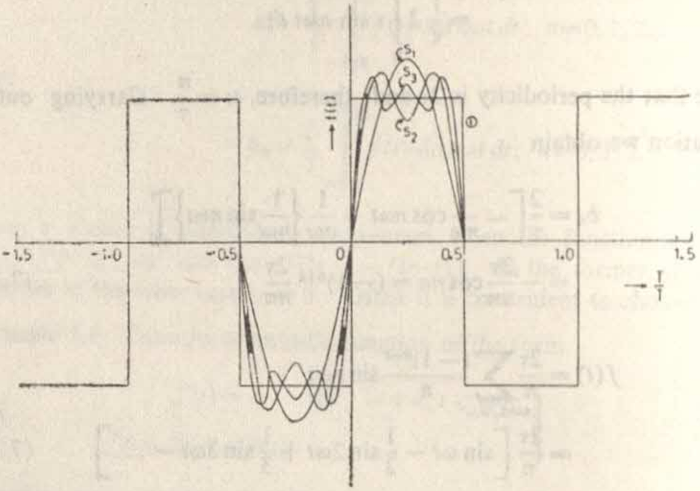


Fig. 7.2 A plot of the periodic function defined by Eq. (7.1-13). S_1, S_2 and S_3 represent the corresponding partial sums.

Thus

$$f(t)=\frac{2A}{\pi}\sum_{n=1,2,3,\ldots}\frac{1}{n}[1-(-1)^n]\sin n\omega t$$

$$=\frac{4A}{\pi}\left[\sin\omega t+\frac{1}{3}\sin 3\omega t+\frac{1}{5}\sin 5\omega t+\ldots\right]$$

The partial sums

$$S_1=\frac{4A}{\pi}\sin\omega t,\; S_2=\frac{4A}{\pi}\left(\sin\omega t+\frac{1}{3}\sin 3\omega t\right)$$

$$S_3=\frac{4A}{\pi}\left[\sin\omega t+\frac{1}{3}\sin 3\omega t+\frac{1}{5}\sin 5\omega t\right]$$

are also plotted in Fig. 7.2.

7.2 Transverse Vibrations of a Plucked String

An interesting application of the Fourier series lies in studying the transverse vibrations of a plucked string.

Let us consider a stretched string, fixed at the two ends. One of the ends (A) is chosen as the origin. In the equilibrium position of the string, it is assumed to lie along the x-axis (see Fig. 7.3). A point of the string is moved upwards by a distance d; the corresponding shape of the string is shown as dashed line in Fig. 7.3. If the displacement occurs at a distance a from the origin, the equation of the string (in its displaced position) would be given by the following equation:

$$\left.\begin{aligned} y &= \frac{d}{a}x && \text{for } 0<x<a \\ &= \frac{d}{L-a}(L-x) && \text{for } a<x<L \end{aligned}\right\} \quad (7.2\text{-}1)$$

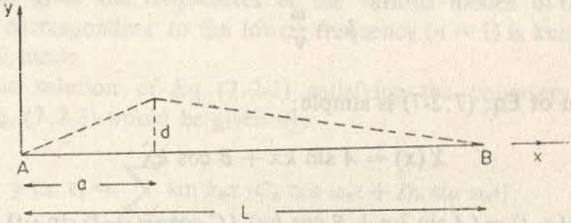


Fig. 7.3 The plucked string; AB represents the equilibrium position. The dashed lines show the displaced position at $t=0$.

where L represents the length of the string. Now, if the string is released from this position at $t=0$, we would like to determine the shape of the string at any subsequent time.

We will show in Sec. 8.6 that the displacement $y(x, t)$ satisfies the following wave equation

$$\frac{\partial^2 y}{\partial x^2} = \frac{1}{v^2}\frac{\partial^2 y}{\partial t^2} \quad (7.2\text{-}2)$$

where v $(=\sqrt{T/\rho})$ represents the speed of the transverse waves, T being the tension in the string and ρ, the mass per unit length. We would like to solve Eq. (7.2-2) subject to the following boundary conditions:

(i) $y=0$ at $x=0$ and $x=L$ for all values of t (7.2-3)

(ii) At $t=0$

(a) $$\frac{\partial y}{\partial t}=0 \quad \text{for all values of } x \quad (7.2\text{-}4)$$

(b) $$\begin{aligned} y(x, t=0) &= \frac{d}{a}x \quad \text{for } 0<x<a \\ &= \frac{d}{L-a}(L-x) \quad \text{for } a<x<L \end{aligned} \quad (7.2\text{-}5)$$

Assuming* a time dependence of the form cos ωt (or sin ωt):

---

*Rigorously we should proceed by using the method of separation of variables; thus we assume

$$y(x, t) = X(x)\,T(t)$$

(*Contd.*)

$$y\,(X,\,t) = X\,(x)\cos\omega t \tag{7.2-6}$$

we obtain

$$\frac{d^2X}{dx^2} = -\frac{\omega^2}{v^2}X\,(x)$$

or,

$$\frac{d^2X}{dx^2} + k^2X\,(x) = 0 \tag{7.2-7}$$

where

$$k = \frac{\omega}{v} \tag{7.2-8}$$

The solution of Eq. (7.2-7) is simple:

$$X(x) = A\sin kx + B\cos kx \tag{7.2-9}$$

Thus

$$y\,(x,\,t) = (A\sin kx + B\cos kx)\;(C\cos\omega t + D\sin\omega t) \tag{7.2-10}$$

Now

$$y\,(x,\,t)\,|_{x=0} = 0 \qquad \text{for all values of } t$$

Thus

$$B = 0$$

Hence

$$y\,(x,\,t) = \sin kx\,[C\sin\omega t + D\sin\omega t] \tag{7.2-11}$$

where we have absorbed A in C and D. Since

$$y\,(x,\,t)\,|_{x=L} = 0 \qquad (\text{for all values of } t)$$

we must have

$$\sin kL = 0$$

or,

$$kL = n\pi;\quad n = 1,\,2,\,3,\ldots \tag{7.2-12}$$

---

where $X\,(x)$ is a function of x alone and $T\,(t)$ is a function of t alone. Substituting in Eq. (7.2-2) we get

$$\frac{1}{X(x)}\frac{d^2X}{dx^2} = \frac{1}{v^2}\frac{1}{T(t)}\frac{d^2T}{dt^2} = -k^2$$

Since the term $\frac{1}{X}\frac{d^2X}{dx^2}$ is a function of x alone and the term $\frac{1}{v^2}\frac{1}{T}\frac{d^2T}{dt^2}$ is a function of t alone, each term must be equal to a constant which we have put equal to $-k^2$. Thus

$$\frac{d^2T}{dt^2} + \omega^2T\,(t) = 0$$

and

$$\frac{d^2X}{dx^2} + k^2X\,(x) = 0$$

where

$$\omega = kv$$

Thus, only discrete values of k (and hence of ω) are permissible; these are given by

$$k_n = \frac{n\pi}{L}; \quad n = 1, 2, \ldots \tag{7.2-13}$$

and

$$\omega_n = \frac{n\pi v}{L}; \; n = 1, 2, \ldots \tag{7.2-14}$$

Eq. (7.2-14) gives the frequencies of the various modes of the system. The mode corresponding to the lowest frequency ($n = 1$) is known as the fundamental mode.

Thus the solution of Eq. (7.2-2) satisfying the boundary condition given by Eq. (7.2-3) would be given by:

$$y(x, t) = \sum_{n=1,2,3,\ldots} \sin k_n x \left[C_n \cos \omega_n t + D_n \sin \omega_n t\right] \tag{7.2-15}$$

Differentiating partially with respect to t we get

$$\left.\frac{\partial y}{\partial t}\right|_{t=0} = \sum_n \sin k_n x \left[-\omega_n C_n \sin \omega_n t + \omega_n D_n \cos \omega_n t\right]\Big|_{t=0}$$

$$= \sum_n \omega_n D_n \sin k_n x \tag{7.2-16}$$

Since

$$\left.\frac{\partial y}{\partial t}\right|_{t=0} = 0 \quad \text{for all values of } x$$

we must have

$$D_n = 0 \quad \text{for all } n$$

Thus

$$y(x, t) = \sum_{n=1,2,3,\ldots} C_n \sin k_n x \cos \omega_n t \tag{7.2-17}$$

or,

$$y(x, 0) = \sum_n C_n \sin\left(\frac{n\pi}{L} x\right) \tag{7.2-18}$$

In order to determine C_n, we employ a method similar to that used in Fourier series. We multiply both sides of Eq. (7.2-18) by $\sin\left(\frac{m\pi}{L} x\right) dx$ and integrate from 0 to L:

$$C_m = \frac{2}{L} \int_0^L y(x, 0) \sin\left(\frac{m\pi}{L} x\right) dx \tag{7.2-19}$$

where we have used the relation

$$\int_0^L \sin\frac{n\pi x}{L}\sin\frac{m\pi x}{L}\,dx=\begin{cases}0 & \text{if } m\neq n\\ L/2 & \text{if } m=n\end{cases} \tag{7.2-20}$$

[cf. Eq. (7.1-5)]. Substituting the expression for $y(x, 0)$ from Eq. (7.2-5), we obtain

$$C_n=\frac{2}{L}\left[\frac{d}{a}\int_0^a x\sin\left(\frac{n\pi x}{L}\right)dx\right.$$

$$\left.+\frac{d}{L-a}\int_a^L (L-x)\sin\left(\frac{n\pi}{L}x\right)dx\right]$$

$$=\frac{2dL^2}{a(L-a)\pi^2 n^2}\sin\left(\frac{n\pi}{L}a\right)$$

On substitution in Eq. (7.2-17), we finally obtain

$$y(x,t)=\frac{2dL^2}{a(L-a)\pi^2}\sum_{n=1,2,3,\ldots}\frac{1}{n^2}\sin\left(\frac{n\pi}{L}a\right)\sin\left(\frac{n\pi}{L}x\right)\cos\left(\frac{n\pi v}{L}t\right) \tag{7.2-21}$$

The above equation can be used to determine the shape of the string at an arbitrary time t. If the string is plucked at the center (i.e. $a = L/2$), terms corresponding to $n=2, 4, 6\ldots$ are absent (i.e. the even harmonics are absent) and Eq. (7.2-21) simplifies to

$$y(x,t)=\frac{8d}{\pi^2}\sum_m (-1)^{m+1}\frac{1}{(2m-1)^2}\sin\frac{(2m-1)\pi x}{L}\cos\frac{(2m-1)(\pi vt)}{L} \tag{7.2-22}$$

The longitudinal vibrations of a rod are discussed in Problem 5.

7.3. Application of Fourier Series in Forced Vibrations

Let us consider the forced vibrations of a damped oscillator. The equation of motion would be

$$m\frac{d^2y}{dt^2}+\Gamma\frac{dy}{dt}+ky=F(t) \tag{7.3-1}$$

where Γ represents the damping constant (see Sec. 6.1) and F represents the external force. It has been shown in Chapter 6 that if $\Gamma>0$ and

$$F(t)=F_0\cos(pt+\theta) \tag{7.3-2}$$

then the steady state solution of Eq. (7.3-1) is a simple harmonic motion with the frequency of the external force. If $F(t)$ is not a sine or cosine function, a general solution of Eq. (7.3-1) is difficult to obtain; however, if $F(t)$ is periodic then we can apply Fourier's theorem to obtain a solution of Eq. (7.3-1). For example, let

$$F(t)=\alpha t \quad \text{for} -\tau<t<\tau \tag{7.3-3}$$

and

$$F(t+2n\tau)=F(t);\ n=1,2,\ldots.$$

The Fourier expansion of such a function was discussed in Example 7.1 and is of the form:

$$F(t)=\frac{2\alpha\tau}{\pi}\left[\sum_{n=1,2,.}\frac{(-1)^{n+1}}{n}\sin n\omega t\right] \tag{7.3-4}$$

We next consider the solution of the differential equation

$$m\,\frac{d^2y_n}{dt^2}+\Gamma\frac{dy_n}{dt}+ky_n=F_n\sin n\omega t$$

or

$$\frac{d^2y_n}{dt^2}+K\frac{dy_n}{dt}+\omega_0^2y_n=A_n\sin n\omega t \tag{7.3-5}$$

where

$$K\equiv\frac{\Gamma}{m};\quad \omega_0^2\equiv\frac{k}{m}$$

and

$$A_n=\frac{F_n}{m}=\frac{(-1)^{n+1}}{n}\,\frac{2\alpha\tau}{\pi m} \tag{7.3-6}$$

The steady state solution of Eq. (7.3-5) will be of the form

$$y_n=C_n\sin n\omega t+D_n\cos n\omega t \tag{7.3-7}$$

and the solution of Eq. (7.3-1) will be of the form

$$y=\sum_n y_n \tag{7.3-8}$$

In order to determine C_n and D_n we substitute the above solution in Eq. (7.3-5) to obtain

$$\begin{aligned}&-n^2\omega^2\,[C_n\sin n\omega t+D_n\cos n\omega t]\\&+n\omega K\,[C_n\cos n\omega t-D_n\sin n\omega t]\\&+\omega_0^2\ \ [C_n\sin n\omega t+D_n\cos n\omega t]=A_n\sin n\omega t\end{aligned}$$

Thus

$$(\omega_0^2-n^2\omega^2)\,C_n-n\omega K\,D_n=A_n$$

and

$$(\omega_0^2-n^2\omega^2)\,D_n+n\omega K\,C_n=0 \tag{7.3-9}$$

Solving the above equations we get

$$D_n = -\frac{n\omega K}{(\omega_0^2 - n^2\omega^2)^2 + n^2\omega^2 K^2} A_n$$

and

$$C_n = \frac{\omega_0^2 - n^2\omega^2}{(\omega_0^2 - n^2\omega^2)^2 + n^2\omega^2 K^2} A_n$$

Thus the steady state solution can be written in the form

$$y = \sum_n G_n \sin(n\omega t + \theta_n) \tag{7.3-10}$$

where the amplitude G_n is given by

$$G_n = (C_n^2 + D_n^2)^{1/2} = \frac{A_n}{[(\omega_0^2 - n^2\omega^2)^2 + n^2\omega^2 K^2]^{1/2}} \tag{7.3-11}$$

7.4 The Fourier Integral

In Sec. 7.1 we had shown that a periodic function can be expanded in the form

$$f(t) = \tfrac{1}{2}a_0 + \sum_{n=1}^{\infty} [a_n \cos n\omega t + b_n \sin n\omega t] \tag{7.4-1}$$

where

$$a_n = \frac{2}{T} \int_{t_0}^{t_0+T} f(t) \cos n\omega t \, dt \tag{7.4-2}$$

$$b_n = \frac{2}{T} \int_{t_0}^{t_0+T} f(t) \sin n\omega t \, dt \tag{7.4-3}$$

and

$$T = \frac{2\pi}{\omega} \tag{7.4-4}$$

On substituting the above expressions for a_n and b_n in Eq. (7.4-1) we get

$$f(t) = \frac{1}{T} \int_{-T/2}^{+T/2} f(t')\, dt' + \sum_{n=1}^{\infty} \left[\frac{2}{T} \cos n\omega t \int_{-T/2}^{+T/2} f(t') \cos n\omega t' \, dt' \right.$$

$$+\frac{2}{T}\sin n\omega t \int_{-T/2}^{+T/2} f(t')\sin n\omega t'\,dt' \Bigg] \qquad (7.4\text{-}5)$$

or

$$f(t)=\frac{1}{2\pi}\,\Delta s \int_{-\pi/\Delta s}^{+\pi/\Delta s} f(t')\,dt' + \sum_{n=1}^{\infty}\frac{\Delta s}{\pi}\int_{-\pi/\Delta s}^{+\pi/\Delta s} f(t')\cos\,[n\Delta s\,(t-t')]\,dt' \qquad (7.4\text{-}6)$$

where

$$\Delta s \equiv \frac{2\pi}{T} = \omega$$

We let $T\rightarrow\infty$ so that* $\Delta s\rightarrow 0$. Thus, if the integral

$$\int_{-\infty}^{+\infty} |\,f(t')\,|\,dt'$$

exists (i.e. if it has a finite value) then the first term on the R.H.S. of Eq. (7.4-6) would go to zero. Further, since

$$\int_{0}^{\infty} F(s)\,ds = \lim_{\Delta S\rightarrow 0}\sum_{n=1}^{\infty} F(n\Delta s)\,\Delta s \qquad (7.4\text{-}7)$$

we have

$$f(t)=\frac{1}{\pi}\int_{0}^{\infty}\Bigg[\int_{-\infty}^{+\infty} f(t')\cos\,[s(t-t')]\,dt'\Bigg]ds \qquad (7.4\text{-}8)$$

Eq. (7.4-8) is known as the Fourier integral. Since the cosine function inside the integral is an even function of s, we may write

$$f(t)=\frac{1}{2\pi}\int_{-\infty}^{+\infty}\Bigg[\int_{-\infty}^{\infty} f(t')\cos\,[s(t-t')]\,dt'\Bigg]ds \qquad (7.4\text{-}9)$$

Further, since $\sin\,[s(t-t')]$ is an odd function of s

$$\frac{i}{2\pi}\int_{-\infty}^{+\infty}\Bigg[\int_{-\infty}^{+\infty} f(t')\sin s(t-t')\,dt'\Bigg]ds=0 \qquad (7.4\text{-}10)$$

---

*When $T\rightarrow\infty$, the function is no more periodic.

Adding Eqs. (7.4-9) and (7.4-10) we get

$$f(t)=\frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(t')\, e^{is(t-t')}\, dt'\, ds \tag{7.4-11}$$

Thus, if

$$F(\omega)\equiv\frac{1}{\sqrt{(2\pi)}}\int_{-\infty}^{+\infty} f(t')\, e^{-i\omega t'}\, dt' \tag{7.4-12}$$

then

$$f(t)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} F(\omega) e^{i\omega t}\, d\omega \tag{7.4-13}$$

where we have replaced s by ω. The function $F(\omega)$ is known as the Fourier transform of $f(t)$ [see also Problem 7.9].

7.5 Group Velocity of a Wave Packet

The displacement corresponding to a one dimensional plane wave propagating in the $+x$ direction can be written in the form

$$\Psi(x, t)=Ae^{i(kx-\omega t)} \tag{7.5.1}$$

where A represents the amplitude* of the wave and other symbols have their usual meaning. However, the plane wave represented by Eq. (7.5-1) is a practical impossibility because at an arbitrary value of t, the displacement is finite for all values of x; for example,

$$\Psi(x, t=0)=Ae^{ikx} \tag{7.5-2}$$

which corresponds to a sinusoidal variation for $-\infty < x < \infty$. In practice, the displacement is finite only over a certain region of space; for example, we may have

$$\begin{aligned} \Psi(x, t=0) &= ae^{ik_0 x} & \quad |x| < \tfrac{1}{2}L \\ &= 0 & \quad |x| > \tfrac{1}{2}L \end{aligned} \tag{7.5-3}$$

Such a wave is known as a wave packet and we will, in this section, study the propagation of such a wave packet.

---

*In general, A may be complex and if we write

$$A=|A|\, e^{i\varphi}$$

then Eq. (7.5-1) becomes

$$\Psi=|A|\, e^{i(kx-\omega t+\varphi)}$$

The actual displacement is the real part of Ψ and is, therefore, given by

$$|A|\cos(kx-\omega t+\varphi)$$

A wave packet can always be expressed as a superposition of plane waves :

$$\Psi(x, t)=\int_{-\infty}^{+\infty} A(k)e^{i(kx-\omega t)}\,dk \tag{7.5-4}$$

where $A(k)$ represents the amplitude of the plane wave corresponding to the wave vector k. At $t=0$ we have

$$\Psi(x, t=0)=\int_{-\infty}^{+\infty} A(k)e^{ikx}\,dk \tag{7.5-5}$$

Thus $\Psi(x, t=0)$ is the Fourier transform of $A(k)$ and inverting the transform, we obtain

$$A(k) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} \Psi(x, t=0)\, e^{-ikx}\,dx \tag{7.5-6}$$

In general, for most cases the function $A(k)$ is very sharply peaked around a certain wave vector k_0. For example, for the wave packet represented by Eq. (7.5-3)

$$A(k)= \frac{1}{2\pi}\int_{-L/2}^{+L/2} a\, e^{i(k_0-k)\,x}\,dx$$

$$=\frac{a}{\pi}\frac{\sin\left[(k-k_0)\,L/2\right]}{(k-k_0)}$$

$$=\frac{a}{\pi k_0}\,\frac{\sin\left[(K-1)\,\alpha\right]}{(K-1)} \tag{7.5-7}$$

where $K=\dfrac{k}{k_0}$ and $\alpha=\dfrac{Lk_0}{2}$. In Fig. 7.4 we have plotted this function for $\alpha=10$. One can see that the function is sharply peaked around $K=1$ (i.e. around $k=k_0$) and the width

$$\Delta K\left(=\frac{\Delta k}{k_0}\right)\sim\frac{1}{\alpha} \tag{7.5-8}$$

Thus

$$\Delta k \sim \frac{1}{L} \tag{7.5-9}$$

It may be noted that for a 1 nanosecond optical pulse (i.e. for $\tau\sim 10^{-9}$ sec. $\lambda\sim 6\times 10^{-5}$ cm)

$$k_0\sim 10^5\ \text{cm}^{-1}$$

and

$$L=c\tau\simeq 30\ \text{cm}\ ,$$

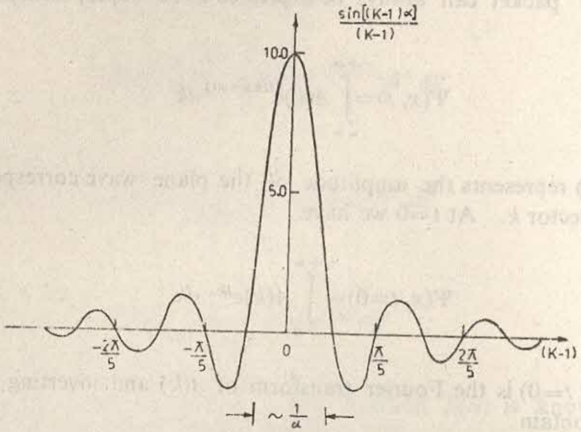


Fig. 7.4 The variation of $\frac{\sin(K-1)\,\alpha}{(K-1)}$ as a function of K for $\alpha=10$.

or

$$\alpha=\frac{k_0 L}{2}\simeq 1.5\times 10^6$$

Hence even for such a short pulse, the value of α is very large and the corresponding spread $\frac{\Delta k}{k_0}$ is extremely small.

Returning to Eq. (7.5-4), we note that since $A(k)$ is sharply peaked around $k=k_0$, we may make a Taylor expansion* of $\omega(k)$ around $k=k_0$:

$$\omega(k)=\omega(k_0)+(k-k_0)\left.\frac{d\omega}{dk}\right|_{k=k_0}+\frac{(k-k_0)^2}{2!}\left.\frac{d^2\omega}{dk^2}\right|_{k=k_0}+\dots \qquad (7.5\text{-}10)$$

We assume that the second and higher order terms could be neglected and hence Eq. (7.5-4) could be written in the form

$$\Psi(x,t)\simeq\int_{-\infty}^{+\infty}A(k)\exp\left[i\left(kx-\omega_0 t-(k-k_0)\left(\frac{d\omega}{dk}\right)_{k=k_0}t\right)\right]dk$$

$$=e^{i(k_0x-\omega_0 t)}\int_{-\infty}^{+\infty}A(k)\exp[i(k-k_0)(x-\mathrm{v}_g t)]\,dk \qquad (7.5\text{-}11)$$

where

$$\mathrm{v}_g\equiv\left.\frac{d\omega}{dk}\right|_{k=k_0} \qquad (7.5\text{-}12)$$

---

*The dependence of $\omega(k)$ on k for optical beams is discussed in Chapter 6.

It should be noted that in the integral appearing on the RHS of Eq. (7.5-11), x and t do not appear independently but only as $(x - \mathrm{v}_g t)$. Thus, if

$$\Psi(x, t=0) = e^{ik_0 x} F(x) \tag{7.5-13}$$

then

$$\Psi(x, t) = e^{i(k_0 x - \omega_0 t)} F(x - \mathrm{v}_g t) \tag{7.5-14}$$

The function $F(x - v_g t)$ represents the envelope of the wave packet and moves with velocity v_g, and therefore v_g is known as the group velocity of the wave packet. For example, corresponding to Eq. (7.5-3), we have

$$F(x) = a \quad |x| < \frac{L}{2}$$

$$= 0 \quad |x| > \frac{L}{2} \tag{7.5-15}$$

The propagation of such a wave packet is shown in Fig. 7.5 and as is obvious, the packet has moved through a distance $\mathrm{v}_g t_0$ in a time t_0.

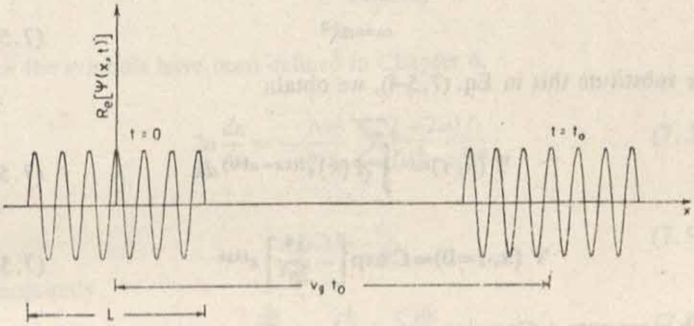


Fig. 7.5 The propagation of a wave packet when second (and higher) order terms are neglected in the expansion of $\omega(k)$ around $k = k_0$. Notice that there is no distortion.

We should however make the following observations;

(i) The quantity

$$\mathrm{v}_p = \frac{\omega}{k} \tag{7.5-16}$$

is known as the phase velocity and has no physical significance when we consider the propagation in a dispersive medium. We may mention here that for a non-dispersive medium

$$\omega \propto k \quad \text{or} \quad \omega = ck$$

where c denotes the speed of propagation of the wave. For such a case, both the group velocity and the phase velocity are equal to c. However, for a dispersive medium, ω is not linearly related to k (see Chapter 6) and it is the group velocity which is of importance.

(ii) As long as we can neglect the second and higher order derivatives (i.e. if we can neglect the term proportional to $\left.\frac{d^2\omega}{dk^2}\right|_{k=k_0}$ etc.), the packet travels without any distortion (See Fig. 7.5). However, if we do take into account these terms, the envelope of the wave packet will change in shape. In fact, it can be shown that the wave packet remains undistorted for $t \ll t_0$, where

$$t_0 \simeq \left| \frac{1}{\beta(\Delta k)^2} \right| \tag{7.5-17}$$

with $\beta = \frac{1}{2}\left.\frac{d^2\omega}{dk^2}\right|_{k=k_0}$. Thus, the time in which a wave packet spreads considerably depends on $\left.\frac{d^2\omega}{dk^2}\right|_{k=k_0}$

Example 7.3 The dispersion relation for a transverse wave in a rigid bar is given by the following equation*

$$\omega = ak^2 \tag{7.5-18}$$

If we substitute this in Eq. (7.5-4), we obtain

$$\Psi(x, t) = \int_{-\infty}^{+\infty} A(k)\, e^{i(kx - ak^2t)}\, dk \tag{7.5-19}$$

Let

$$\Psi(x, t=0) = C \exp\left[-\frac{x^2}{2\sigma^2}\right] e^{ik_0x} \tag{7.5-20}$$

which represents a Gaussian wave packet. Thus

$$A(k) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} C \exp\left[-\frac{x^2}{2\sigma^2} + ik_0x - ikx\right] dx \tag{7.5-21}$$

Carrying out the integration, we obtain

$$A(k) = \frac{C}{2\pi}\sqrt{2\sigma^2}\sqrt{\pi}\exp\left[-\frac{(k-k_0)^2\sigma^2}{2}\right] \tag{7.5-22}$$

---

*This follows from the fact that in a rigid bar, the displacement of a transverse wave satisfies the following equation

$$\frac{\partial^2 y}{\partial t^2} = \alpha \frac{\partial^4 y}{\partial x^4}$$

Thus if we substitute a solution of the form exp $[i(kx - \omega t)]$ one immediately obtains the dispersion relation.

which shows that $A(k)$ is peaked around $k=k_0$ with $\Delta k \sim \frac{1}{\sigma}$. Thus

$$\Psi(x,t)=\frac{Ca}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}\exp\left[-\frac{(k-k_0)^2}{2}\sigma^2+ikx-iak^2t\right]dk$$

$$=\frac{C\sigma}{\sqrt{2\pi}}\frac{\sqrt{\pi}}{\left(\frac{\sigma^2}{2}+iat\right)^{1/2}}\exp\left[\frac{(k_0\sigma^2+ix)^2}{4\left(\frac{\sigma^2}{2}+iat\right)}-\frac{k_0^2\sigma^2}{2}\right] \tag{7.5-23}$$

If we simplify the above expression and take its real part (see Problem 10) we will see that the wave packet remains Gaussian whose width increases with time. Indeed, the width doubles itself in times of the order of σ^2/a which is consistent with Eq. (7.5-17)—see Fig. 7.6.

Example 7.4 In Chapter 6 we have shown that the dependence of the refractive index of a dielectric on frequency is of the form:

$$n^2(\omega)=1+\frac{Ne^2}{m\epsilon_0}\sum_j\frac{f_j}{\omega_j^2-\omega^2} \tag{7.5-24}$$

where the symbols have been defined in Chapter 6.
Thus

$$2n\frac{dn}{d\omega}=-\frac{Ne^2}{m\epsilon_0}\sum_j\frac{(-2\omega)f_j}{(\omega_j^2-\omega^2)^2} \tag{7.5-25}$$

Now

$$n(\omega)=\frac{c}{\omega/k} \tag{7.5-26}$$

Consequently

$$\frac{dn}{d\omega}=-\frac{ck}{\omega^2}+\frac{c}{\omega}\frac{dk}{d\omega} \tag{7.5-27}$$

or

$$v_g=\frac{d\omega}{dk}=\frac{c}{n+\omega\frac{dn}{d\omega}} \tag{7.5-28}$$

Substituting for $\frac{dn}{d\omega}$ from Eq. (7.5-25), we get the group velocity. For $\omega \gg \omega_j$ (i.e. for high frequencies)

$$v_g \approx \frac{c\left[1-\frac{Ne^2}{2m\epsilon_0}\frac{1}{\omega^2}\right]}{\left[1+\frac{Ne^2}{m\epsilon_0}\left(-\frac{1}{\omega^2}\right)\right]+\left[\frac{Ne^2}{m\epsilon_0}\left(\frac{1}{\omega^2}\right)\right]}$$

$$\approx c\left[1-\frac{Ne^2}{2m\epsilon_0}\frac{1}{\omega^2}\right] \tag{7·5-29}$$

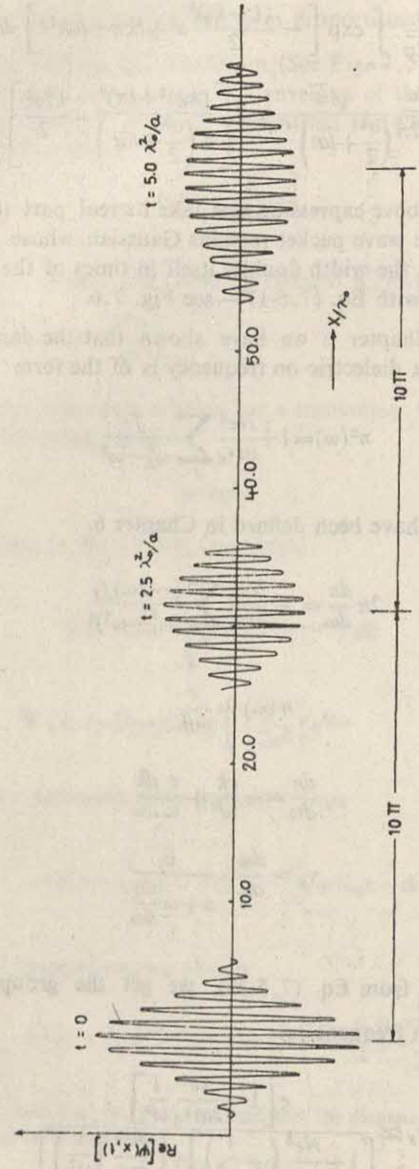


Fig. 7.6 The broadening of a Gaussian pulse in a dispersive medium characterised by Eq. (7.5-18).

Similarly, one can calculate v_g for $\omega \ll \omega_j \ldots$*

Example 7.5 For an electron wave one has

$$E = \frac{p^2}{2m} \tag{7.5-30}$$

where E represents the energy, p the momentum of the electron and m the mass of the electron. If we now use the relation (see Chapter 21)

$$E = \hbar\omega \tag{7.5-31}$$

and

$$p = \hbar k \tag{7.5-32}$$

one obtains the dispersion relation

$$\omega = \frac{\hbar k^2}{2m} \tag{7.5-33}$$

Thus

$$\frac{d^2\omega}{dk^2} = \frac{\hbar}{m} \tag{7.5-34}$$

For an electron $m \simeq 10^{-27}$ g and one has

$$\frac{d^2\omega}{dk^2} \simeq 1 \text{ cm}^2/\text{sec}$$

Thus, if an electron is localised within a distance $L \sim 10^{-8}$ cm (i.e. $\triangle k \simeq 10^8 \text{ cm}^{-1}$), then

$$t_0 \simeq 10^{-16} \text{ sec}$$

where we have used Eq. (7.5-17). Thus the wave packet associated with an electron spreads considerably with time. On the other hand, for a dust particle with $m \cong 10^{-5}$ g and $L \sim 10^{-3}$ cm one gets

$$t_0 \cong 2 \times 10^{16} \text{ sec} \sim 10^9 \text{ years}$$

which shows that a classical description is adequate.

---

*For ω close to the resonant frequency, the present analysis is not applicable because the attenuation can not be neglected—and k will become complex (see Chapter 6)

PROBLEMS

1. Consider a periodic force of the form:

$$F(t)=F_0 \sin \omega t \quad \text{for} \quad 0<t<T/2$$
$$=0 \quad \text{for} \quad T/2<t<T$$

and $$F(t+T)=F(t);$$

where

$$\omega=2\pi/T$$

Show that

$$F(t)=\tfrac{1}{2} F_0+\tfrac{1}{2} F_0 \sin \omega t-2/\pi\, F_0\,(1/3 \cos 2\omega t+1/15 \cos 4\omega t+\ldots)$$

One obtains a periodic voltage of the above form in a half wave rectifier. What will be the Fourier expansion corresponding to full wave rectification?

2. For the periodic force discussed in Problem 1, solve Eq. (7.3-1) and obtain the steady state solution.

3. Consider a time dependent electric field of the form

$$E(t)=E_0 e^{i\omega_0 t} \qquad -\tau<t<\tau$$
$$=0 \qquad |t|>\tau$$

Obtain its Fourier transform and interpret the result physically for $\omega_0\tau \gg 1$.

4. Consider an electric field with Gaussian envelope:

$$E(t)=E_0 \exp\left[-\frac{t^2}{\tau^2}\right] \exp\,(i\omega_0 t)$$

Show that its Fourier transform is also Gaussian. Interpret the result physically for $\omega_0\tau \gg 1$.

5. Consider longitudinal waves propagating in a rod (of length L) whose one end is fixed and the other end is free. Show that the displacement would be given by

$$\xi(x,t)=\sum_{n=0,1,2} A_n \sin\left(\frac{\omega_n}{v}x\right)\cos(\omega_n t+\varphi_n)$$

where $\omega_n=\dfrac{(2n+1)\pi c}{2L}$, c being the speed of the longitudinal waves and A_n are constants.

(Hint: Assume the fixed end to be $x=0$. The other end being free implies that $\dfrac{\partial\xi}{\partial x}$ at $x=L$ is zero at all times; physically, this is because of the fact that there is no stress at $x=L$ and the stress is proportional to $\dfrac{\partial\xi}{\partial x}$).

6. If v_g and v_p represent the group and phase velocities, show that they are related through the following equation

$$v_p-\lambda\frac{dv_p}{d\lambda}=v_g$$

7. The deep water waves are characterized by the following dispersion relation

$$\omega=c\sqrt{k}$$

Show that the group velocity is half of the phase velocity.

8. If we consider the propagation of transverse elastic waves in a rod, the phase velocity can be shown to be proportional to the wave vector. For such a case show that the group velocity is twice of the phase velocity. Show also that one obtains the same dispersion relation for the deBroglie waves associated with a free electron.

9. (a) The Dirac delta function, $\delta(x-a)$, is defined through the following equations:

and $$\delta(x-a)=0 \quad \text{for} \quad x \neq a$$

$$\int_{a-\alpha}^{a+\alpha} f(x)\,\delta(x-a)=f(a)$$

where α is an arbitrary positive quantity and $f(x)$ is a 'well-behaved' function. There are many possible representations of the delta function. Show that

$$\operatorname*{Lt}_{g\to\infty} \frac{\sin g\,(x-x')}{\pi\,(x-x')}=\delta\,(x-x')$$

(b) Since

$$\frac{\sin g\,(x-x')}{\pi\,(x-x')}=\frac{1}{2\pi}\int_{-g}^{+g} e^{ik\,(x-x')}\,dk,$$

one obtains another representation of the delta function

$$\frac{1}{2\pi}\int_{-\infty}^{\infty} e^{ik\,(x-x')}\,dk=\delta\,(x-x')$$

Using the above representation prove Eqs. (7.4-11)—(7.4-13).

10. In Example 7.3 show that

$$\text{Re}[\Psi\,(x,t)]=\frac{C}{\left[1+\frac{\zeta^2}{\eta^4}\right]^{1/4}} \exp\left[-\frac{\eta^2(\xi-\zeta)}{2\,(\zeta^2+\eta^4)}\right]\cos\left[\frac{\eta^4\,(2\xi-\zeta)+\xi^2\zeta}{2\,(\zeta^2+\eta^4)}-\frac{1}{2}\tan^{-1}\frac{\zeta}{\eta^2}\right]$$

where

$$\xi=k_0x,\ \eta=k_0\sigma \text{ and } \zeta=2\,k_0^2\,at.$$

REFERENCES AND SUGGESTED READING

Bottcher, C.J.F. (1952), *Theory of Electric Polarization*, Elsevier Publishing Co., Amsterdam (Netherlands).

Braddick, H.J.J. (1965), *Vibrations, Waves and Diffraction*, McGraw-Hill Publishing Co., London (England).

Crawford, F.S. (1968), *Waves and Oscillations: Berkeley Physics Course*, Vol. III, McGraw-Hill Book Co., New York (USA).

Feynman, R.P., Leighton. R.B. and **Sands, M.** (1965), *The Feynman Lectures on Physics*, Vol. I, Addison-Wesley Publishing Co., Reading, Mass. (USA).

French, A.P. (1973), *Vibrations and Waves*, Arnold-Heinman India, New Delhi.

Loudon, R. (1973), *The Quantum Theory of Light*, Clarendon Press, Oxford (England).

Pain, H.J. (1968), *The Physics of Vibrations and Waves*, John Wiley & Sons, London (England).

Resnik, R. and **Halliday, D.** (1966), *Physics*, Part I, John Wiley & Sons, New York (USA).

Sommerfeld, A. (1964). *Optics*, Academic Press, New York (USA).

Stone, J.M. (1963), *Radiation and Optics*, McGraw-Hill Book Co., New York (USA).

8 Wave Propagation and the Wave Equation

> *Ten thousand saw I at a glance,*
> *Tossing their heads in sprightly dance*
> *The waves beside them danced; but they*
> *out-did the sparkling waves in glee:*
> *I gazed and gazed but little thought*
> *What wealth the show to me had brought.*
>
> WORDSWORTH

8.1 Introduction

In this chapter we will discuss the phenomenon of waves. A wave is propagation of a disturbance. For example, when we drop a small stone in a calm pool of water, a circular pattern spreads out from the point of impact. The impact of the stone creates a disturbance which propagates outwards. In this propagation, the water molecules do not move outward with the wave; instead they move in nearly circular orbits about an equilibrium position (see Sec. 8.3). Once the disturbance has passed a certain region, every drop of water is left at its original position. This fact can easily be verified by placing a small piece of wood on the surface of water. As the wave passes, the piece of wood makes oscillations and once the disturbance has passed, the wood comes back to its original position. Further, with time the circular ripples spread out, i.e. the disturbance (which is confined to particular region at a given time) produces a similar disturbance at a neighbouring point at a slightly later time with the pattern of disturbance roughly remaining the same. Such propagation of disturbances (without any translation of the medium in the direction of propagation) is termed as a wave. It is also seen that the wave carries energy; in this case the energy is in the form of kinetic energy of water molecules.

We will first consider the simplest example of wave propagation, viz. the propagation of a transverse wave on a string. Consider yourself holding one end of a string, the other end being, held tightly by another person so that the string does not sag. If you move the end of the string up and down a few times then a disturbance is created which propagates towards the other end of the string. Thus, if we take a snapshot of the string at $t=0$ and at a slightly later time Δt, then the snapshots will roughly* look like the ones shown in Figs. 8.1 (a) and (b). The figures show that the disturbances have identical shapes except for the fact that one is displaced from the other by a distance $v\Delta t$ where v represents the speed of the disturbance. Such a propagation of a disturbance without its change in form is characteristic of a wave. The following points may, however, be noted:

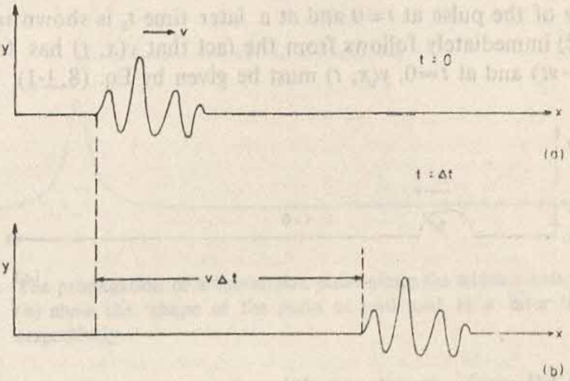


Fig. 8.1 A transverse wave is propagating along the $+x$-axis on a string; (a) and (b) show the displacements at $t=0$ and $t=\Delta t$ respectively.

(i) A certain amount of work is done when the wave is generated and as the wave propagates through the string, it carries with it a certain amount of energy which is felt by the person holding the other end of the string.

(ii) The wave is transverse, i.e. the displacement of the particles of the string is at right angles to the direction of propagation of the wave.

Referring back to Figs. 8.1 (a) and (b), we note that the shape of the string at the instant Δt is similar to its shape at $t=0$, except for the fact that the whole disturbance has travelled through a certain distance. If v represents the speed of the wave then this distance is simply $v\Delta t$. Consequently, if the equation describing the rope at $t=0$ is $y(x)$ then at a later instant t, the equation of the curve would be $y(x-vt)$ which simply implies a shift of the origin by a distance vt. Similarly, for a disturbance propagating in the $-x$ direction, if the equation describing the rope at $t=0$ is $y(x)$ then at a later instant t the equation of the curve would be $y(x+vt)$.

Example 8.1 Study the propagation of a semicircular pulse in the $+x$ direction whose displacement at $t=0$ is given by the following equations:

---

*We are assuming here that as the disturbance propagates through the string, there is negligible attenuation and also no change in the shape of the disturbance.

$$\begin{aligned} y(x, t=0) &= +[R^2-x^2]^{1/2} \qquad &|x| \le R \\ &= 0 \qquad &|x| \ge R \end{aligned} \tag{8.1-1}$$

Solution. For a wave propagating in the $+x$ direction the dependence of $y(x, t)$ on x and t should be through the function $(x-vt)$. Consequently,

$$\begin{aligned} y(x, t) &= +[R^2-(x-vt)^2]^{1/2} \qquad &|x-vt| \le R \\ &= 0 \qquad &|x-vt| \ge R \end{aligned} \tag{8.1-2}$$

The shape of the pulse at $t=0$ and at a later time t_0 is shown in Fig. 8.2. Eq. (8.1-2) immediately follows from the fact that $y(x, t)$ has to be of the form $y(x-vt)$ and at $t=0$, $y(x, t)$ must be given by Eq. (8.1-1).

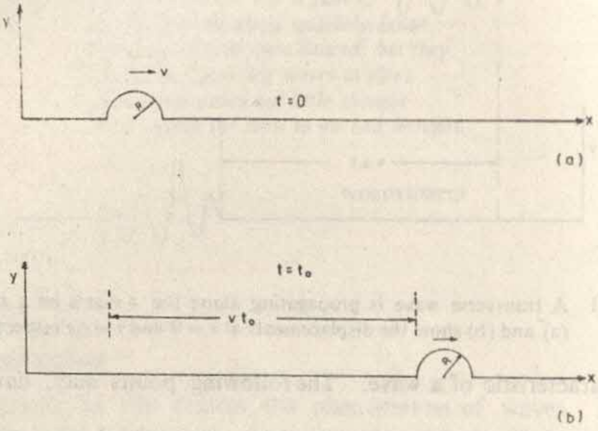


Fig. 8.2 The propagation of a semicircular pulse along the $+x$-axis; (a) and (b) show the shape of the pulse at $t=0$ and at a later time t_0 respectively.

Example 8.2 Consider a pulse propagating in the minus x direction with speed v. The shape of the pulse at $t=t_0$ is given by

$$y(x, t=t_0) = \frac{b^2}{a^2+(x-x_0)^2} \tag{8.1-3}$$

(Such a pulse is known as a Lorentzian pulse). Determine the shape of the pulse at an arbitrary time t.

Solution. The shape of the pulse at $t=t_0$ is shown in Fig. 8.3 (a). The maximum of the displacement occurs at $x=x_0$. Since the pulse is propagating in the $-x$ direction, at a later time t, the maximum will occur at $x_0-\mathrm{v}(t-t_0)$. Consequently, the shape of the pulse at an arbitrary time t would be given by

$$y(x, t) = \frac{b^2}{a^2+[x-x_0+\mathrm{v}\,(t-t_0)]^2} \tag{8.1-4}$$

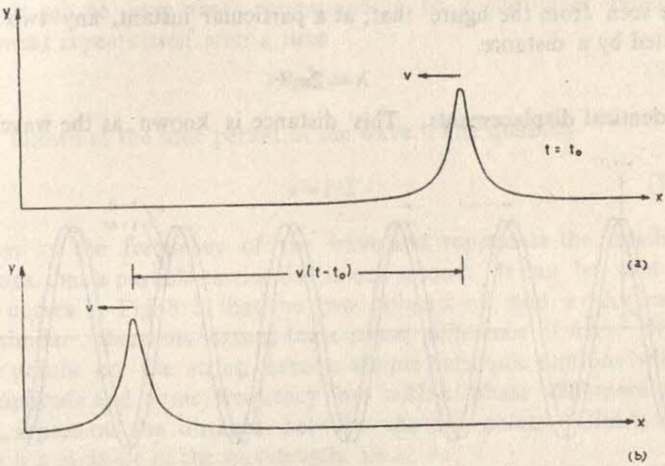


Fig. 8.3 The propagation of a Lorentzian pulse along the minus x-axis; (a) and (b) show the shape of the pulse at $t=0$ and at a later instant t_0 respectively.

Eq. (8.1-4) could have been written down directly from Eq. (8.1-3) by replacing x by $x+v(t-t_0)$.

8.2 Sinusoidal Waves : Concept of Frequency and Wavelength

Till now we have been considering the propagation of a pulse which lasts for a finite amount of time. We will now consider a periodic wave in which the displacement $y(x, t)$ has the form

$$y(x, t) = a \cos k\,(x \mp vt + \varphi) \tag{8.2-1}$$

where the upper and lower signs correspond to waves propagating in the $+x$ and $-x$ directions respectively. Such a displacement is indeed produced in a long stretched string at the end of which a continuously vibrating tuning fork is placed. The quantity φ is known as the phase of the wave (see Chapter 5). We may, without loss of generality, assume $\varphi=0$. Thus for a wave propagating in the $+x$ direction

$$y(x, t) = a \cos k\,(x - vt) \tag{8.2-2}$$

In Fig. 8.4 we have plotted the dependence of the displacement y on x at $t=0$ and at $t=\Delta t$. These are given by

$$y(x) = a \cos kx \qquad \text{at} \quad t=0$$

and

$$y(x) = a \cos k\,(x - v\Delta t) \qquad \text{at} \quad t=\Delta t$$

(8 2-3

The two curves are the snapshots of the string at the two instants. It can be seen from the figure that, at a particular instant, any two points separated by a distance

$$\lambda = 2\pi/k \tag{8.2-4}$$

have identical displacements. This distance is known as the wavelength.

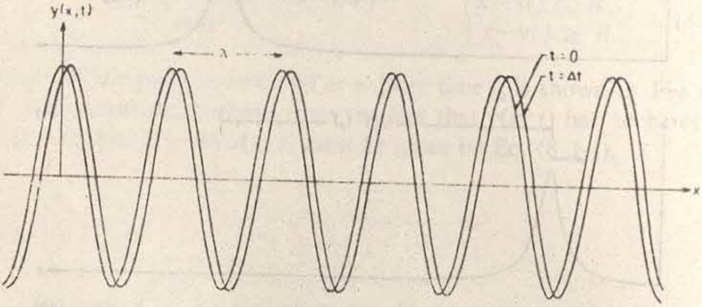


Fig. 8.4 The curves represent the displacement of a string at $t = 0$ and at $t = \Delta t$ respectively when a sinusoidal wave is propagating in the $+x$-direction.

Further, the displaced curve (which corresponds to the instant $t = \Delta t$) can be obtained by displacing the curve corresponding to $t = 0$ by a distance $v\Delta t$; this shows that the wave is propagating in the $+x$ direction with speed v. It can also be seen that the maximum displacement of the particle (from its equilibrium position) is a, which is known as the amplitude of the wave.

In Fig. 8.5 we have plotted the time dependence of the displacement of the points characterised by $x = 0$ and $x = \Delta x$. These are given by

$$\left.\begin{aligned} y(t) &= a \cos \omega t \qquad &&\text{at } x = 0 \\ y(t) &= a \cos (\omega t - k\Delta x) \qquad &&\text{at } x = \Delta x \end{aligned}\right\} \tag{8.2-5}$$

and

where

$$\omega = kv \tag{8.2-6}$$

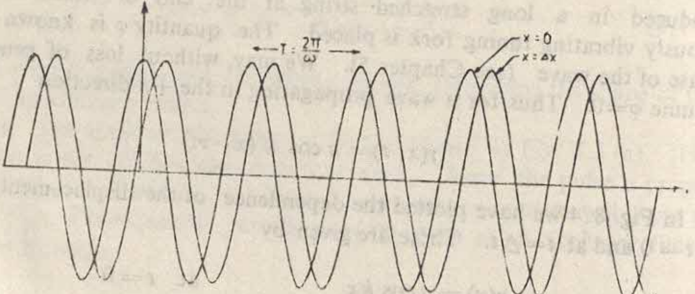


Fig. 8.5 The curves represent the time variation of the displacement at $x = 0$ and at $x = \Delta x$ respectively when a sinusoidal wave is propagating in the $+x$-direction.

The curves correspond to the time variation of the displacement of the two points. It can be seen that, corresponding to a particular point, the displacement repeats itself after a time

$$T=2\pi/\omega \tag{8.2-7}$$

which is known as the time period of the wave. The quantity

$$\nu=1/T \tag{8.2-8}$$

is known as the frequency of the wave and represents the number of oscillations that a particle carries out in one second. It can be seen from the two curves in Fig. 8.5, that the two points $x=0$ and $x=\Delta x$ execute exactly similar vibrations except for a phase difference of $k\Delta x$. In fact any two points on the string execute simple harmonic motions with the same amplitude and same frequency but with a phase difference of kx_0 where x_0 represents the distance between the two points. Clearly if this distance is a multiple of the wavelength, i.e.

$$x_0 = m\lambda,\ m = 1, 2, \ldots$$

then

$$kx_o = \frac{2\pi}{\lambda}\, m\lambda = 2m\pi$$

which implies that two points separated by a distance which is a multiple of the wavelength vibrate with the same phase. Similarly, two points separated by a distance $\frac{1}{2}\lambda$, $\frac{3}{2}\lambda$ . . . vibrate in opposite phase. In general, a path difference of x_0 corresponds to a phase difference of $\frac{2\pi}{\lambda}\, x_0$.

Using Eqs. (8.2-6)–(8.2-8) we get

$$\nu = \frac{1}{T} = \frac{\omega}{2\pi} = \frac{kv}{2\pi} = \frac{v}{\lambda}$$

or

$$v = \nu\lambda \tag{8.2-9}$$

Notice the similarity in the variation of the displacement with respect to x (at a given value of time) and with respect to t (at a given value of x); see Figs. 8.4 and 8.5. The similarity can be expressed by writing Eq. (8.2-2) in the form

$$y(x, t) = a \cos\left(\frac{2\pi}{\lambda}\, x - \frac{2\pi}{T}\, t\right) \tag{8.2-10}$$

which shows that the wavelength λ in Fig. 8.4 plays the same role as the time period T in Fig. 8.5. Eq. (8.2-10) is often written in the form

$$y(x, t) = a \cos (kx-\omega t) \tag{8 2-11}$$

It should be pointed out that the entire discussion given above would remain valid for an arbitrary value of the phase factor φ.

8.3 Types of Waves

As mentioned earlier, when a wave is propagating through a string the displacement is at right angles to the direction of propagation. Such a wave is known as a transverse wave.* Similarly, when a sound wave propagates through air the displacement of the air molecules are along the direction of propagation of the wave; such waves are known as longitudinal waves. However, there are waves which are neither longitudinal nor transverse in character; for example, when a wave propagates through the surface of water, the water molecules move approximately in a circular orbit.

8.4 Energy Transport in Wave Motion

A wave carries energy; for example, when a transverse wave propagates through a string, the particles execute simple harmonic motions about their equilibrium position and associated with this motion is a certain amount of energy. As the wave propagates through, the energy gets transported from one end of the string to the other. We consider the time variation of the displacement of a particle which can be written as

$$y = a \cos(\omega t + \varphi) \tag{8.4-1}$$

The instantaneous velocity of the particle would be

$$v = \frac{dy}{dt} = -a\omega \sin(\omega t + \varphi) \tag{8.4-2}$$

Thus, the kinetic energy (T) would be given by

$$\begin{aligned} T &= \tfrac{1}{2} m \left(\frac{dy}{dt}\right)^2 \\ &= \tfrac{1}{2} m a^2 \omega^2 \sin^2(\omega t + \varphi) \end{aligned} \tag{8.4-3}$$

The total energy (E) will be the maximum value of T

$$\begin{aligned} E &= (T)_{\max} \\ &= \tfrac{1}{2} m a^2 \omega^2 [\sin^2(\omega t + \varphi)]_{\max} \\ &= \tfrac{1}{2} m a^2 \omega^2 \end{aligned} \tag{8.4-4}$$

For a sound wave propagating through a gas, the energy per unit volume, ϵ, would be given by

$$\begin{aligned} \epsilon &= \tfrac{1}{2} m n a^2 \omega^2 \\ &= \tfrac{1}{2} \rho a^2 \omega^2 \\ &= 2\pi^2 \rho a^2 \nu^2 \end{aligned} \tag{8.4-5}$$

---

*Electromagnetic waves are also transverse in character. However, it should be mentioned that the electromagnetic waves have also a longitudinal component near the source which dies off rapidly at large distances (see Sec. 19.4).

where n represents the number of molecules per unit volume and $\rho (=nm)$ the density of the gas. With such a wave, we can associate the intensity which is defined as the energy flow per unit time across a unit area perpendicular to the direction of propagation. Since the speed of propagation of the wave is v, the intensity (I) would be given by*

$$I = 2\pi^2 \rho v a^2 \nu^2 \tag{8.4-6}$$

Thus the intensity is proportional to the square of the amplitude and square of the frequency.

Let us consider a wave emanating from a point source in a uniform isotropic** medium. We will assume that there is no absorption and that the source is emitting W joules per second (W represents the power of the source). Consider a sphere of radius r whose centre is at the point source. Clearly, W joules per second will cross the spherical surface whose area is $4\pi r^2$. Thus, the intensity (I) would be given by

$$I = \frac{W}{4\pi r^2} \tag{8.4-7}$$

which is nothing but the inverse square law. Using Eqs. (8.4-6) and (8.4-7) we obtain

$$\frac{W}{4\pi r^2} = 2\pi^2 \rho v a^2 \nu^2$$

or

$$a = \left[\frac{W}{8\pi^3 \rho v \nu^2}\right]^{1/2} \frac{1}{r} \tag{8.4-8}$$

showing that the amplitude falls off as $\frac{1}{r}$. Indeed, for a spherical wave*** emanating from a point source, the displacement is given by

$$f = \frac{a_0}{r} \sin (kr - \omega t)$$

where a_0 represents the amplitude of the wave at an unit distance from the source.

---

*This can be easily understood from the fact that if we have N particles per unit volume each moving with the same velocity v then the number of particles crossing an unit area (normal to v) per unit time would be Nv.

**Isotropic substances are the ones in which physical properties (like velocity of propagation of a particular wave) are the same in all directions. In chapter 18 we will consider anisotropic substances.

***When waves emanate from a point source in an isotropic medium, all the points on the surface of a sphere (whose centre is at the point source) have the same amplitude and the same phase; in other words, the locus of points which have the same amplitude and the same phase is a sphere. Such waves are known as spherical waves. Far away from the source, over a small area, the spherical waves are essentially plane waves.

Example 8.3 A source of sound is vibrating with a frequency of 256 vibrations per second in air and propagating energy uniformly in all directions at the rate of 5 joules per second. Calculate the intensity and the amplitude of the wave at a distance of 25 m from the source. Assume that there is no absorption [speed of sound waves in air=330 m/sec; density of air=1·29 kg/m^3].

Solution.

$$\text{Intensity} \quad I=\frac{5\text{ joules/sec}}{4\pi\times(25)^2\text{ m}^2}\approx 6.3\times10^{-4}\text{ J sec}^{-1}\text{ m}^{-2}$$

$$\text{Thus} \quad a=\left[\frac{5}{8\pi^3\times1.29\times330\times256\times256}\right]^{1/2}\frac{1}{25}$$

$$\approx 3\times10^{-7}\text{ m.}$$

Example 8.4 Show that when a transverse wave propagates through a string, the energy transmitted per unit time is $\frac{1}{2}\rho\omega^2a^2\text{v}$ where ρ is the mass per unit length, a the amplitude of the wave and v the speed of propagation of the wave.

Solution. The energy associated per unit length of the string is $\frac{1}{2}\rho\omega^2a^2$ since the speed of the wave is v the result follows.

8.5 The One-Dimensional Wave Equation

In Sec. 8.1 we had shown that the displacement ψ of a one-dimensional wave is always of the form

$$\psi=f(x-\text{v}t)+g(x+\text{v}t) \tag{8.5-1}$$

where the first term on the RHS of the above equation represents a disturbance propagating in the $+x$ direction with speed v and similarly, the second term represents a disturbance propagating in the $-x$ direction with speed v. The question now arises as to how can we predict the existence of waves and what would be the velocity of propagation of these waves? The answer to this question is as follows: If we can derive an equation of the form

$$\frac{\partial^2\psi}{\partial x^2}=\frac{1}{\text{v}^2}\frac{\partial^2\psi}{\partial t^2} \tag{8.5-2}$$

from physical considerations then we can be sure that waves will result and ψ will represent the displacement associated with the wave. This follows from the fact that the general solution of Eq. (8.5-2) is of the form

$$\psi=f(x-\text{v}t)+g(x+\text{v}t) \tag{8.5-3}$$

where f and g are arbitrary functions of their argument. Although the general solution of Eq. (8.5-2) will be derived in Sec. 8.9, we will show here that the solution expressed by Eq. (8.5-3) indeed satisfies Eq. (8.5-2). We rewrite Eq. (8.5-3) in the form

$$\psi = f(\xi) + g(\eta) \tag{8.5-4}$$

where

$$\xi = x - \mathrm{v}t \tag{8.5-5a}$$

and

$$\eta = x + \mathrm{v}t \tag{8.5-5b}$$

Thus

$$\frac{\partial \psi}{\partial x} = f'(\xi)\frac{\partial \xi}{\partial x} + g'(\eta)\frac{\partial \eta}{\partial x}$$

where primes denote differentiation with respect to the argument, i.e.

$$f'(\xi) = \frac{df}{d\xi}, \text{ etc.}$$

Since $\dfrac{\partial \xi}{\partial x} = 1 = \dfrac{\partial \eta}{\partial x}$ [see Eq. (8.5-5)], we obtain

$$\frac{\partial \psi}{\partial x} = f'(\xi) + g'(\eta)$$

Differentiating once more, we get

$$\begin{aligned} \frac{\partial^2 \psi}{\partial x^2} &= f''(\xi)\frac{\partial \xi}{\partial x} + g''(\eta)\frac{\partial \eta}{\partial x} \\ &= f''(\xi) + g''(\eta) \end{aligned} \tag{8.5-6}$$

Similarly,

$$\begin{aligned} \frac{\partial \psi}{\partial t} &= f'(\xi)\frac{\partial \xi}{\partial t} + g'(\eta)\frac{\partial \eta}{\partial t} \\ &= \mathrm{v}[-f'(\xi) + g'(\eta)] \end{aligned}$$

and

$$\frac{\partial^2 \psi}{\partial t^2} = \mathrm{v}^2[f''(\xi) + g''(\eta)] \tag{8.5-7}$$

Thus

$$\frac{1}{\mathrm{v}^2}\frac{\partial^2 \psi}{\partial t^2} = f''(\xi) + g''(\eta) = \frac{\partial^2 \psi}{\partial x^2} \tag{8.5-8}$$

which proves that the solution expressed by Eq. (8.5-3) indeed satisfies Eq. (8.5-2). Consequently, if we ever obtain an equation of the form of Eq. (8.5-2) from physical considerations, we can predict the existence of waves, the speed of which would be v.

We must mention that the simplest particular solutions of the wave equation correspond to sinusoidal variation:

$$\psi = A \sin[k(x \pm \mathrm{v}t) + \varphi], \tag{8.5-9a}$$

or

$$\psi = A \cos[k(x \pm \mathrm{v}t) + \varphi]. \tag{8.5-9b}$$

As shown in Sec. 8.2

$$k = 2\pi/\lambda \quad \text{and} \quad kv = \omega = 2\pi\nu$$

where λ is the wavelength and ν the frequency of the wave. Instead of sinusoidal variation it is often more convenient to write the solution in the form

$$\psi = A \exp[i(kx \pm \omega t + \varphi)], \tag{8.5-10}$$

where, as before, A and φ represent the amplitude and phase of the wave. In writing Eq. (8.5-10), it is implied that the actual displacement is just the real part of ψ which is

$$A \cos(kx \pm \omega t + \varphi).$$

In the next three sections we will derive the wave equation for some simhle cases.* In Sec. 8.9 we will discuss the general solution of the wave equation.

8.6 Transverse Vibrations of a Stretched String

Let us consider a stretched string having a tension T. In its equilibrium position the string is assumed to lie on the x-axis. If the string is pulled in the y-direction then forces will act on the string which will tend to bring it back to its equilibrium position. Let us consider a small length AB of the string and calculate the net force acting on it in the y-direction. Due to the tension T, the end points A and B experience force in the direction of the arrows shown in Fig. 8.6. The force at A in the upward direction is

$$-T\sin\theta_1 \approx -T\tan\theta_1 = -T\frac{\partial y}{\partial x}\bigg|_x$$

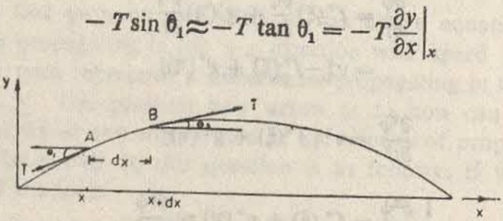


Fig. 8.6 Transverse vibrations of a stretched string.

Similarly, the force at B in the upward direction is

$$T\sin\theta_2 \approx T\tan\theta_2 = T\frac{\partial y}{\partial x}\bigg|_{x+dx}$$

where we have assumed θ_1 and θ_2 to be small. Thus the net force acting on AB in the y-direction is

---

*In Chapter 19 we will derive the wave equation from Maxwell's equations and thereby obtain an expression for the speed of electromagnetic waves.

$$T\left[\left(\frac{\partial y}{\partial x}\right)_{x+dx}-\left(\frac{\partial y}{\partial x}\right)_x\right]$$

$$=T\frac{\partial^2 y}{\partial x^2}dx, \qquad (8.6\text{-}1)$$

where we have used the Taylor series expansion of $\left(\frac{\partial y}{\partial x}\right)_{x+dx}$ about the point x:

$$\left(\frac{\partial y}{\partial x}\right)_{x+dx}=\left(\frac{\partial y}{\partial x}\right)_x+\frac{\partial}{\partial x}\left(\frac{\partial y}{\partial x}\right)\bigg|_x dx$$

and have neglected higher order terms because dx is infinitesimal. The equation of motion is therefore

$$\Delta m\,\frac{\partial^2 y}{\partial t^2}=T\,\frac{\partial^2 y}{\partial x^2}dx$$

where Δm is the mass of the element AB. If ρ is the mass per unit length then

$$\Delta m=\rho\,dx$$

and we get

$$\frac{\partial^2 y}{\partial x^2}=\frac{1}{T/\rho}\frac{\partial^2 y}{\partial t^2} \qquad (8.6\text{-}2)$$

which is the one-dimensional wave equation. Thus we may conclude that transverse waves can propagate through a stretched string and if we compare the above equation with Eq. (8.5-2) we obtain the following expression for the speed of the transverse waves:

$$v=\sqrt{T/\rho}. \qquad (8.6\text{-}3)$$

The vibrations of a clamped string will be discussed in Sec 10.2. It should be mentioned that in an actual string, the displacement is not rigorously of the form given by Eq. (8.5-1); there is, in general, an attenuation of the wave and also the shape does not remain unaltered. This is a consequence of the various approximations made in the derivation of the wave equation.

8.7 Longitudinal Sound Waves in a Solid

In this section we will derive an expression for the velocity of longitudinal sound waves propagating in an elastic solid. Let us consider a solid cylindrical rod of cross-sectional area A. Let PQ and RS be two transverse sections of the rod at distances x and $x+\Delta x$ from a fixed point O, where we have chosen the x-axis to be along the length of the rod (see Fig. 8.7).

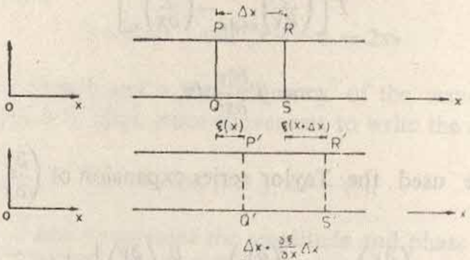


Fig. 8.7 Propagation of longitudinal sound waves through a cylindrical rod.

Let the longitudinal displacement of a plane be denoted by $\xi(x)$. Thus the displacements of the planes *PQ* and *RS* would be $\xi(x)$ and $\xi(x+\Delta x)$ respectively. In the displaced position, the distance between the planes *P′Q′* and *R′S′* would be

$$\xi(x+\Delta x)-\xi(x)+\Delta x$$

$$=\xi(x)+\frac{\partial\xi}{\partial x}\Delta x-\xi(x)+\Delta x$$

$$=\Delta x+\frac{\partial\xi}{\partial x}\Delta x$$

The elongation of the element would be $\frac{\partial\xi}{\partial x}\Delta x$ and, therefore, the longitudinal strain would be

$$\frac{\text{Increase in length}}{\text{Original length}}=\frac{\frac{\partial\xi}{\partial x}\Delta x}{\Delta x}=\frac{\partial\xi}{\partial x} \qquad (8.7\text{-}1)$$

Since the Young's modulus (Y) is defined as the ratio of the longitudinal stress to the longitudinal strain, we have

$$\text{Longitudinal stress}=\frac{F}{A}=Y\times\text{Strain}$$

$$=Y\frac{\partial\xi}{\partial x} \qquad (8.7\text{-}2)$$

where F is the force acting on the element *P′Q′*. Thus

$$F(x)=YA\frac{\partial\xi}{\partial x} \qquad (8.7\text{-}3)$$

and, therefore,

$$\frac{\partial F}{\partial x}=YA\frac{\partial^2\xi}{\partial x^2} \qquad (8.7\text{-}4)$$

Now, if we consider the volume $P'Q'S'R'$ then a force F is acting on the element $P'Q'$ in the negative x-direction and a force $F(x+\Delta x)$ is acting on the plane $R'S'$ along the positive x-direction. Thus the resultant force acting on the element $P'Q'S'R'$ will be

$$F(x+\Delta x)-F(x) = \frac{\partial F}{\partial x}\,\Delta x$$

$$= YA\,\frac{\partial^2 \xi}{\partial x^2}\,\Delta x \tag{8.7-5}$$

If ρ represents the density, then the mass of the element would be $\rho A\,\Delta x$. Thus the equation of motion will be

$$\rho A\,\Delta x\,\frac{\partial^2 \xi}{\partial t^2} = YA\,\Delta x\,\frac{\partial^2 \xi}{\partial x^2}$$

or

$$\frac{\partial^2 \xi}{\partial x^2} = \frac{1}{v_l^2}\,\frac{\partial^2 \xi}{\partial t^2} \tag{8.7-6}$$

where

$$v_l = \sqrt{\frac{Y}{\rho}} \tag{8.7-7}$$

represents the velocity of the waves and the subscript l refers to the fact that we are considering longitudinal waves.*

The above derivation is valid when the transverse dimension of the rod is small compared with the wavelength of the disturbance so that one may assume that the longitudinal displacement at all points on any transverse section (like PQ) are the same. In general, if one carries out a rigorous analysis of the vibrations of an extended isotropic elastic solid, one can show that the velocities of the longitudinal and transverse waves will be given by**

$$v_l = \sqrt{\frac{Y}{\rho}\,\frac{(1-\sigma)}{(1+\sigma)(1-2\sigma)}} = \sqrt{\frac{K+\frac{4}{3}\eta}{\rho}} \tag{8.7-8}$$

$$v_t = \sqrt{\frac{Y}{\rho}\,\frac{1}{2(1+\sigma)}} = \sqrt{\frac{\eta}{\rho}} \tag{8.7-9}$$

where σ, η and K represent the Poisson ratio, modulus of rigidity and bulk modulus respectively. We must mention that the transverse wave

---

*In a similar manner one can consider transverse waves propagating through an elastic solid, the velocity of which would be given by [see, for example, Slater and Frank (1933)],

$$v_t = \sqrt{\eta/\rho}$$

where η represents the modulus of rigidity.

**See, for example, Joos (1955).

[whose velocity is given by Eq. (8.7-9)] is due to the restoring forces arising because of the elastic properties of the material, whereas corresponding to the transverse waves discussed in Sec. 8.6, the string moved as a whole and the restoring force was due to the externally applied tension.

8.8 Longitudinal Waves in a Gas

In order to determine the speed of propagation of longitudinal sound waves in a gas, we consider a column *PQSR* as shown in Fig. 8.8 (a).

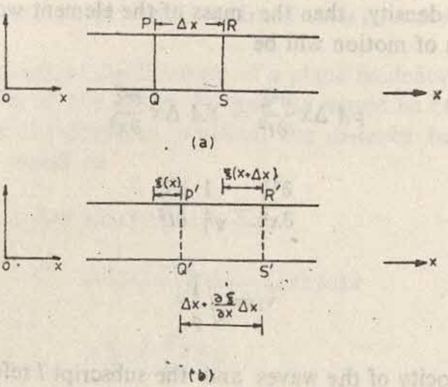


Fig. 8.8 Propagation of longitudinal sound waves through air.

Once again, because of a longitudinal displacement, the plane *PQ* gets displaced by $\xi(x)$ and the plane *RS* gets displaced by a distance $\xi(x+\Delta x)$ (see Fig. 8.8). Let the pressure of the gas in the absence of any disturbance be P_0. Let $P_0+\Delta P(x)$ and $P_0+\Delta P(x+\Delta x)$ denote the pressures at the planes *P′Q′* and *R′S′* respectively. Now, if we consider the column *P′Q′S′R′* then the pressure $P_0+\Delta P(x)$ on the face *P′Q′* acts in the $+x$ direction whereas the pressure $P_0+\Delta P(x+\Delta x)$ on the face *R′S′* acts in the $-x$ direction. Thus the force acting on the column *P′Q′S′R′* would be

$$[\Delta P(x)-\Delta P(x+\Delta x)]\,A$$

$$= -\frac{\partial}{\partial x}(\Delta P)\,\Delta x\,A \tag{8.8-1}$$

where *A* represents the cross-sectional area. Consequently, the equation of motion for the column *P′Q′S′R′* would be

$$-\frac{\partial}{\partial x}(\Delta P)\,A\,\Delta x = \rho A\,\Delta x\,\frac{\partial^2\xi}{\partial t^2}$$

where ρ represents the density of the gas. Thus

$$-\frac{\partial}{\partial x}(\Delta P) = \rho\,\frac{\partial^2\xi}{\partial t^2} \tag{8.8-2}$$

Now, a change in pressure gives rise to a change in volume, and if the frequency of the wave is large ($\geq 20Hz$), the pressure fluctuations will be rapid and one may assume the process to be adiabatic. Thus, we may write

$$PV^{\gamma} = \text{constant} \tag{8.8-3}$$

where $\gamma = C_p/C_v$ represents the ratio of the two specific heats. If we differentiate the above expression, we get

$$\Delta P\, V^{\gamma} + \gamma V^{\gamma-1} P \cdot \Delta V = 0$$

or

$$\Delta P = -\frac{\gamma P}{V}\,\Delta V, \tag{8.8-4}$$

The change in the length of the column *PQSR* is

$$[\xi(x+\Delta x) - \xi(x) + \Delta x] - \Delta x$$

$$= \frac{\partial \xi}{\partial x}\,\Delta x$$

Thus, the change in the volume

$$\Delta V = \frac{\partial \xi}{\partial x}\,A\,\Delta x$$

The original volume V of the element is $A\,\Delta x$. Thus

$$\Delta P = -\frac{\gamma P}{A\,\Delta x}\,\frac{\partial \xi}{\partial x}\,A\,\Delta x$$

$$= -\gamma P\,\frac{\partial \xi}{\partial x}, \tag{8.8-5}$$

or

$$\frac{\partial}{\partial x}(\Delta P) = -\gamma P\,\frac{\partial^2 \xi}{\partial x^2}. \tag{8.8-6}$$

Using Eqs. (8.8-2) and (8.8-6) we obtain

$$\frac{\partial^2 \xi}{\partial x^2} = \frac{1}{v^2}\,\frac{\partial^2 \xi}{\partial t^2} \tag{8.8-7}$$

where

$$v = \sqrt{\frac{\gamma P}{\rho}} \tag{8.8-8}$$

represents the velocity of propagation of longitudinal sound waves in a gas. For air, if we assume $\gamma = 1.40$, $P = 1.01 \times 10^6$ dynes/cm^2 and $\rho = 1.3 \times 10^{-3}$ g/cm^3, then we obtain

$$v \approx 330 \text{ m/sec}$$

It is of interest to mention that the adiabatic compressibility of a gas is given by

$$\kappa_s = -\frac{1}{V}\left(\frac{\partial V}{\partial P}\right)_s = \frac{1}{\gamma P} \tag{8.8·9}$$

where the subscript s refers to the adiabatic condition (constant entropy). The bulk modulus (K) of a gas is the inverse of κ_S:

$$K = \frac{1}{\kappa_s} = \gamma P \tag{8.8-10}$$

and if we substitute this expression for K in Eq. (8.7-8), we obtain Eq. (8.8-8) where we have used the fact that the modulus of rigidity (η) for a gas is zero.

8.9. The General Solution of the One Dimensional Wave Equation*

In order to obtain a general solution of the equation

$$\frac{\partial^2\psi}{\partial x^2} = \frac{1}{v^2}\frac{\partial^2\psi}{\partial t^2} \tag{8.9-1}$$

we introduce two new variables

$$\xi = x - vt \tag{8.9-2}$$

$$\eta = x + vt \tag{8.9-3}$$

and write Eq. (8.9-1) in terms of these variables. Now,

$$\frac{\partial\psi}{\partial x} = \frac{\partial\psi}{\partial\xi}\frac{\partial\xi}{\partial x} + \frac{\partial\psi}{\partial\eta}\frac{\partial\eta}{\partial x} \tag{8.9.4}$$

or

$$\frac{\partial\psi}{\partial x} = \frac{\partial\psi}{\partial\xi} + \frac{\partial\psi}{\partial\eta} \tag{8.9-5}$$

where we have used the fact that

$$\frac{\partial\xi}{\partial x} = 1 \text{ and } \frac{\partial\eta}{\partial x} = 1$$

Differentiating Eq. (8.9-5) with respect to x we get

$$\frac{\partial^2\psi}{\partial x^2} = \frac{\partial}{\partial x}\left(\frac{\partial\psi}{\partial\xi}\right) + \frac{\partial}{\partial x}\left(\frac{\partial\psi}{\partial\eta}\right)$$

---

*May be skipped in the first reading.

$$= \frac{\partial}{\partial \xi}\left(\frac{\partial \psi}{\partial \xi}\right)\frac{\partial \xi}{\partial x} + \frac{\partial}{\partial \eta}\left(\frac{\partial \psi}{\partial \xi}\right)\frac{\partial \eta}{\partial x} + \frac{\partial}{\partial \xi}\left(\frac{\partial \psi}{\partial \eta}\right)\frac{\partial \xi}{\partial x} + \frac{\partial}{\partial \eta}\left(\frac{\partial \psi}{\partial \eta}\right)\frac{\partial \eta}{\partial x}$$

or

$$\frac{\partial^2 \psi}{\partial x^2} = \frac{\partial^2 \psi}{\partial \xi^2} + 2\frac{\partial^2 \psi}{\partial \eta \partial \xi} + \frac{\partial^2 \psi}{\partial \eta^2} \qquad (8.9\text{-}6)$$

Similarly

$$\frac{\partial \psi}{\partial t} = \frac{\partial \psi}{\partial \xi}\frac{\partial \xi}{\partial t} + \frac{\partial \psi}{\partial \eta}\frac{\partial \eta}{\partial t}$$
$$= -v\frac{\partial \psi}{\partial \xi} + v\frac{\partial \psi}{\partial \eta}$$

and

$$\frac{\partial^2 \psi}{\partial t^2} = -v\left[\frac{\partial}{\partial \xi}\left(\frac{\partial \psi}{\partial \xi}\right)\frac{\partial \xi}{\partial t} + \frac{\partial}{\partial \eta}\left(\frac{\partial \psi}{\partial \xi}\right)\frac{\partial \eta}{\partial t}\right] + v\left[\frac{\partial}{\partial \xi}\left(\frac{\partial \psi}{\partial \eta}\right)\frac{\partial \xi}{\partial t} + \frac{\partial}{\partial \eta}\left(\frac{\partial \psi}{\partial \eta}\right)\frac{\partial \eta}{\partial t}\right]$$

or

$$\frac{\partial^2 \psi}{\partial t^2} = v^2\left[\frac{\partial^2 \psi}{\partial \xi^2} - 2\frac{\partial^2 \psi}{\partial \eta \partial \xi} + \frac{\partial^2 \psi}{\partial \eta^2}\right] \qquad (8.9\text{-}7)$$

Substituting the expressions for $\partial^2\psi/\partial x^2$ and $\partial^2\psi/\partial t^2$ from Eqs. (8.9-6) and (8.9-7) in Eq. (8.9-1) we obtain

$$\frac{\partial^2 \psi}{\partial \xi^2} + 2\frac{\partial^2 \psi}{\partial \eta \partial \xi} + \frac{\partial^2 \psi}{\partial \eta^2} = \frac{\partial^2 \psi}{\partial \xi^2} - 2\frac{\partial^2 \psi}{\partial \eta \partial \xi} + \frac{\partial^2 \psi}{\partial \eta^2}$$

or

$$\frac{\partial}{\partial \eta}\left(\frac{\partial \psi}{\partial \xi}\right) = 0 \qquad (8.9\text{-}8)$$

Thus $\partial\psi/\partial\xi$ has to be independent of η; however, it can be an arbitrary function of ξ:

$$\frac{\partial \psi}{\partial \xi} = F(\xi) \qquad (8.9\text{-}9)$$

or

$$\psi = \int F(\xi)\, d\xi + \text{constant of integration.}$$

The constant of integration can be an arbitrary function of η and since the integral of an arbitrary function is again an arbitrary function, we obtain the following general solution of the wave equation

$$\psi = f(\xi) + g(\eta)$$
$$= f(x - vt) + g(x + vt), \qquad (8.9\text{-}10)$$

where f and g are arbitrary functions of their argument. The function

$f(x-vt)$ represents a disturbance propagating in the $+x$ direction with speed v and the function $g(x+vt)$ represents a disturbance propagating in the $-x$ direction.

Example 8.5· Solve the one-dimensional wave equation [Eq. (8.5-2)] by the method of separation of variables[*] and show that the solution can indeed be expressed in the form given by Eqs. (8.5-9) and (8.5-10).

Solution. In the method of separation of variables, we try a solution of the wave equation

$$\frac{\partial^2\psi}{\partial x^2}=\frac{1}{v^2}\frac{\partial^2\psi}{\partial t^2} \tag{8.9-11}$$

of the form

$$\psi(x,t)=X(x)\,T(t) \tag{8.9-12}$$

where $X(x)$ is a function of x alone and $T(t)$ is a function of t alone. Substituting in Eq. (8.9-11), we get

$$T(t)\frac{d^2X}{dx^2}=\frac{1}{v^2}X(x)\frac{d^2T}{dt^2}$$

or[**]

$$\frac{1}{X(x)}\frac{d^2X}{dx^2}=\frac{1}{v^2T(t)}\frac{d^2T}{dt^2} \tag{8.9-13}$$

The LHS is a function of x alone and the RHS is a function of t alone. This implies that a function of one independent variable x is equal to a function of another independent variable t for all values of x and t. This is possible only when each side is equal to a constant; we set this constant equal to $-k^2$, thus

$$\frac{1}{X(x)}\frac{d^2X}{dx^2}=\frac{1}{v^2}\frac{1}{T(t)}\frac{d^2T}{dt^2}=-k^2 \tag{8.9-14}$$

or

$$\frac{d^2\partial}{dx^2}+k^2X(x)=0 \tag{8.9-15}$$

and

$$\frac{d^2T}{dt^2}+\omega^2T(t)=0 \tag{8.9-16}$$

where

$$\omega=kv=\frac{2\pi v}{\lambda} \tag{8.9-17}$$

---

[*] The method of separation of variables is a powerful method for solving certain kinds of partial differential equations. According to this method the solution is assumed to be a product of functions, each function depending only on one independent variable (see Eq. 8.9-12). On substituting this solution, if the variables separate out then the method is said to work and the general solution is a linear sum of all possible solutions. If the variables do not separate out one has to try some other method to solve the equation.

[**] Notice that partial derivatives have been replaced by total derivatives.

represents the angular frequency of the wave. The solutions of Eqs. (8.9-15) and (8.9-16) can easily be written down:

$$X(x) = (A\cos kx + B\sin kx) \tag{8.9-18}$$

and

$$T(t) = (C\cos\omega t + D\sin\omega t)$$

Thus

$$\psi(x, t) = (A\cos kx + B\sin kx)(C\cos\omega t + D\sin\omega t)$$

Suitable choice of the constants A, B, C and D would give

$$\psi(x, t) = a\cos(kx - \omega t + \varphi)$$

or

$$\psi(x, t) = a\cos(kx + \omega t + \varphi)$$

representing waves propagating in the $+x$ and $-x$ directions respectively. One can also have

$$\psi(x, t) = a\exp[\pm i(kx \pm \omega t + \varphi)]$$

as a solution.

In general, all values of the frequencies are possible, but the frequency and wavelength have to be related through Eq. (8.9-16). However, there are systems (like a string under tension and fixed at both ends) where only certain values of frequencies are possible (see Sec. 7.2).

Example 8.6 Till now we have confined our discussion to waves in one-dimension. The three-dimensional wave equation is of the form

$$\nabla^2\psi = \frac{1}{v^2}\frac{\partial^2\psi}{\partial t^2} \tag{8.9-19}$$

where

$$\nabla^2\psi \equiv \frac{\partial^2\psi}{\partial x^2} + \frac{\partial^2\psi}{\partial y^2} + \frac{\partial^2\psi}{\partial z^2} \tag{8.9-20}$$

Solve the three-dimensional wave equation by the method of separation of variables and interpret the solution physically.

Solution. Using the method of separation of variables, we write

$$\psi(x, y, z, t) = X(x)\, Y(y)\, Z(z)\, T(t) \tag{8.9-21}$$

where $X(x)$ is a function of x alone, etc. Substituting in Eq. (8.9-19) we obtain

$$YZT\frac{d^2X}{dx^2} + XZT\frac{d^2Y}{dy^2} + XYT\frac{d^2Z}{dz^2} = \frac{1}{v^2}XYZ\frac{d^2T}{dt^2}$$

or dividing through by ψ

$$\left[\frac{1}{X}\frac{d^2X}{dx^2}\right] + \left[\frac{1}{Y}\frac{d^2Y}{dy^2}\right] + \left[\frac{1}{Z}\frac{d^2Z}{dz^2}\right] = \frac{1}{v^2}\left[\frac{1}{T}\frac{d^2T}{dt^2}\right] \tag{8.9-22}$$

Since the first term on the LHS is a function of x alone, the second term is a function of y alone, etc. each term must be set equal to a constant. We write

$$\frac{1}{X}\frac{d^2X}{dx^2} = -k_x^2$$

$$\frac{1}{Y}\frac{d^2Y}{dy^2} = -k_y^2 \qquad (8.9\text{-}23)$$

$$\frac{1}{Z}\frac{d^2Z}{dz^2} = -k_z^2$$

where k_x^2, k_y^2 and k_z^2 are constants. Thus

$$\frac{1}{v^2}\left[\frac{1}{T}\frac{d^2T}{dt^2}\right] = -(k_x^2+k_y^2+k_z^2)$$

or

$$\frac{d^2T}{dt^2} + \omega^2 T(t) = 0 \qquad (8.9\text{-}24)$$

where

$$\omega^2 = k^2 v^2 \qquad (8.9\text{-}25)$$

and

$$k^2 = k_x^2 + k_y^2 + k_z^2$$

The solutions of Eqs. (8.9-23) and (8.9-24) could be written in terms of sine and cosine functions; however, it is more convenient to write them in terms of the exponentials:

$$\begin{aligned}\psi &= A\exp[i(k_x x + k_y y + k_z z \pm \omega t + \varphi)]\\ &= A\exp[i(\mathbf{k}.\mathbf{r} \pm \omega t + \varphi)]\end{aligned} \qquad (8.9\text{-}26)$$

where the vector **k** is defined such that its x, y and z-components are k_x, k_y and k_z respectively. One could have also written

$$\psi = A\cos(\mathbf{k}\cdot\mathbf{r} - \omega t + \varphi) \qquad (8.9\text{-}27)$$

Consider a vector **r** which is normal to **k**; thus $\mathbf{k}\cdot\mathbf{r}=0$; consequently at a given time the phase of the disturbance is constant on a plane normal to **k**. The direction of propagation of the disturbance is along **k** and the phase fronts are planes normal to **k**; such waves are known as plane waves (see Fig. 8.9). Notice that for a given value of the frequency, the value of k^2 is fixed (see Eq. 8.9-25); however, we can have waves propagating in different directions depending on the values of k_x, k_y and k_z. For example, if

$$k_x = k \quad \text{and} \quad k_y = k_z = 0 \qquad (8.9\text{-}28a)$$

we have a wave propagating along the x-axis, the phase fronts are parallel to the y-z plane. Similarly, for

$$k_x = \frac{k}{\sqrt{2}},\ k_y = \frac{k}{\sqrt{2}},\ k_z = 0 \tag{8.9-28b}$$

the waves are propagating in a direction which makes equal angles with x and y-axes, etc. (see Fig. 8.9).

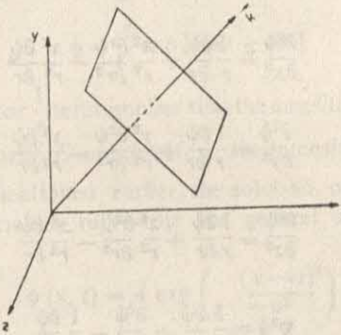


Fig. 8.9 Propagation of a plane wave along the direction **k**.
$\left(k_x = k_y = \frac{k}{\sqrt{2}},\ k_z = 0\right)$

Example 8.7 For a spherical wave the displacement ψ depends only on r and t where r is the magnitude of the distance from a fixed point. Obtain a general solution of the wave equation for a spherical wave.
Solution. We will first show that for a spherical wave

$$\nabla^2\psi = \frac{\partial^2\psi}{\partial r^2} + \frac{2}{r}\frac{\partial\psi}{\partial r} = \frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial\psi}{\partial r}\right) \tag{8.9-29}$$

Now*

$$\frac{\partial\psi}{\partial x} = \frac{\partial\psi}{\partial r}\frac{\partial r}{\partial x} \tag{8.9-30}$$

Since

$$r^2 = x^2 + y^2 + z^2 \tag{8.9-31}$$

therefore,

$$2r\,\frac{\partial r}{\partial x} = 2x$$

or

$$\frac{\partial r}{\partial x} = \frac{x}{r}$$

---

*In general ψ is a function of the three spherical polar coordinates r, θ and φ and

$$\frac{\partial\psi}{\partial x} = \frac{\partial\psi}{\partial r}\frac{\partial r}{\partial x} + \frac{\partial\psi}{\partial\theta}\frac{\partial\theta}{\partial x} + \frac{\partial\psi}{\partial\varphi}\frac{\partial\varphi}{\partial x}$$

However, for a spherical wave ψ is a function of r alone and the last two terms vanish.

Thus

$$\frac{\partial\psi}{\partial x}=\frac{x}{r}\frac{\partial\psi}{\partial r}$$

and

$$\frac{\partial^2\psi}{\partial x^2}=\frac{1}{r}\frac{\partial\psi}{\partial r}+x\frac{\partial}{\partial r}\left(\frac{1}{r}\frac{\partial\psi}{\partial r}\right)\frac{\partial r}{\partial x}$$

$$=\frac{1}{r}\frac{\partial\psi}{\partial r}+\frac{x^2}{r}\left[\frac{1}{r}\frac{\partial^2\psi}{\partial r^2}-\frac{1}{r^2}\frac{\partial\psi}{\partial r}\right]$$

or

$$\frac{\partial^2\psi}{\partial x^2}=\frac{1}{r}\frac{\partial\psi}{\partial r}+\frac{x^2}{r^2}\frac{\partial^2\psi}{\partial r^2}-\frac{x^2}{r^3}\frac{\partial\psi}{\partial r}$$

Similarly

$$\frac{\partial^2\psi}{\partial y^2}=\frac{1}{r}\frac{\partial\psi}{\partial r}+\frac{y^2}{r^2}\frac{\partial^2\psi}{\partial r^2}-\frac{y^2}{r^3}\frac{\partial\psi}{\partial r}$$

and

$$\frac{\partial^2\psi}{\partial z^2}=\frac{1}{r}\frac{\partial\psi}{\partial r}+\frac{z^2}{r^2}\frac{\partial^2\psi}{\partial r^2}-\frac{z^2}{r^3}\frac{\partial\psi}{\partial r}$$

Adding, we get

$$\nabla^2\psi=\frac{3}{r}\frac{\partial\psi}{\partial r}+\frac{\partial^2\psi}{\partial r^2}-\frac{1}{r}\frac{\partial\psi}{\partial r}$$

where we have used Eq. (8.9-31). Thus

$$\nabla^2\psi=\frac{\partial^2\psi}{\partial r^2}+\frac{2}{r}\frac{\partial\psi}{\partial r}=\frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial\psi}{\partial r}\right) \tag{8.9-32}$$

Thus, the wave equation for a spherical wave simplifies to

$$\nabla^2\psi=\frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial\psi}{\partial r}\right)=\frac{1}{v^2}\frac{\partial^2\psi}{\partial t^2} \tag{8.9-33}$$

If we make the substitution

$$\psi=\frac{u(r,t)}{r}\ ,$$

then

$$\frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial\psi}{\partial r}\right)=\frac{1}{r^2}\frac{\partial}{\partial r}\left(r\frac{\partial u}{\partial r}-u\right)$$

$$=\frac{1}{r}\frac{\partial^2 u}{\partial r^2}$$

Thus Eq. (8.9-33) becomes

$$\frac{1}{r}\frac{\partial^2 u}{\partial r^2}=\frac{1}{v^2}\frac{1}{r}\frac{\partial^2 u}{\partial t^2}$$

or

$$\frac{\partial^2 u}{\partial r^2}=\frac{1}{v^2}\frac{\partial^2 u}{\partial t^2} \tag{8.9-34}$$

which is of the same form as the one dimensional wave equation. The general solution of Eq. (8.9-34) is therefore given by

$$\psi=\frac{f(r-\mathrm{v}t)}{r}+\frac{g(r+\mathrm{v}t)}{r} \tag{8.9-35}$$

the first and the second terms (on the RHS) representing an outgoing spherical wave and an incoming spherical wave respectively. For time dependence of the form exp ($\pm\, i\omega t$) one obtains

$$\psi=\frac{A}{r}\exp\left[i\left(kr\pm\omega t\right)\right] \tag{8.9-36}$$

Notice that the factor $\frac{1}{r}$ term implies that the amplitude of a spherical wave decreases inversely with r, and therefore, the intensity will fall off as $1/r^2$.

Example 8.8 As mentioned earlier, the solution obtained by the method of separation of variables represents the general solution. Show that a pulse of the form

$$\psi(x,t)=A\exp\left(\frac{(x-\mathrm{v}t)^2}{-\sigma^2}\right) \tag{8.9-37}$$

can be expressed as a superposition of the solutions obtained by the method of separation of variables.

Solution. An arbitrary disturbance propagating in the $+x$ direction can be written as a superposition of the functions

$$\cos k(x-\mathrm{v}t) \text{ and } \sin k(x-\mathrm{v}t)$$

Thus

$$\psi(x,t)=\int_0^\infty a(k)\cos k(x-\mathrm{v}t)\,dk+\int_0^\infty b(k)\sin k(x-\mathrm{v}t)\,dk \tag{8.9-38}$$

where the functions $a(k)$ and $b(k)$ are to be determined from the form of $\psi(x, t=0)$. Now,

$$\psi(x,0)=\int_0^\infty a(k)\cos kx\,dk+\int_0^\infty b(k)\sin kx\,dk \tag{8.9-39}$$

If $\psi(x, 0)$ is an even function of x, $b(k)=0$ and if $\psi(x,0)$ is an odd function of x, $a(k)=0$. (If $\psi(x, 0)$ is neither even nor odd, both $a(k)$ and $b(k)$ will be finite.) In the present case

$$\psi(x,0)=A\exp\left(-\frac{x^2}{\sigma^2}\right) \tag{8.9-40}$$

which is an even function of x. Thus

$$\psi(x,0)=A\exp\left(-\frac{x^2}{\sigma^2}\right)=\int_0^\infty a(k)\cos kx\,dk \tag{8.9-41}$$

Using the Fourier cosine transform

$$a(k) = \frac{2}{\pi} \int_0^\infty \psi(x, 0) \cos kx \, dx$$

$$= \frac{2}{\pi} A \int_0^\infty \exp\left(-\frac{x^2}{\sigma^2}\right) \cos kx \, dx$$

$$= \frac{\sigma}{\sqrt{\pi}} A \exp\left(-\frac{k^2\sigma^2}{4}\right)$$

Hence

$$A \exp\left[-\left(\frac{x - vt}{\sigma^2}\right)^2\right] = \frac{\sigma A}{\sqrt{\pi}} \int_0^\infty \exp\left(-\frac{k^2\sigma^2}{4}\right) \cos (kx - \omega t) \, dk \qquad (8.9\text{-}42)$$

PROBLEMS

1. The displacement associated with a wave is given by

   (i) $y(x, t) = 0.1 \cos (0.2x - 2t)$

   (ii) $y(x, t) = 0.2 \sin (0.5x + 3t)$

   (iii) $y(x, t) = 0.5 \sin 2\pi (0.1x - t)$

   where in each case x and y are measured in centimeters and t in seconds. Calculate the wavelength, amplitude, frequency and the velocity in each case.
2. A transverse wave ($\lambda = 15$ cm, $\nu = 200$ sec^{-1}) is propagating on a stretched string in the $+x$-direction with an amplitude of 0.5 cm. At $t = 0$ the point $x = 0$ is at its equilibrium position moving in the upward direction. Write the equation describing the wave and if $\rho = 0.1$ g/cm, calculate the energy associated with the wave per unit length of the wire.
3. Assuming that the human ear can hear in the frequency range $20 < \nu < 20{,}000$ Hz, what will be the corresponding wavelength range?
4. Calculate the speed of longitudinal waves at NTP in (a) argon ($\gamma = 1.33$) (b) Hydrogen ($\gamma = 1.66$). (Ans: 380 m/sec. 169 m/sec)
5. Consider a wave propagating in the $+x$-direction with speed 100 cm/sec. The displacement at $x = 10$ cm is given by the following equation

   $$y(x = 10, t) = 0.5 \sin (0.4t)$$

   where x and y are measured in centimeters and t in seconds. Calculate the wavelength and the frequency associated with the wave and obtain an expression for the time variation of the displacement at $x = 0$.
6. Consider a wave propagating in the $-x$-direction whose frequency is 100 sec^{-1}. At $t = 5$ sec the displacement associated with the wave is given by the following equation:

   $$y(x, t = 5) = 0.5 \cos (0.1\,x)$$

   where x and y are measured in centimeters and t in seconds. Obtain the displacement (as a function of x) at $t = 10$ sec. What is the wavelength and the velocity associated with the wave?
7. Repeat the above problem corresponding to

   $$y(x, t = 5) = 0.5 \cos (0.1\,x) + 0.4 \sin (0.1\,x + \pi/3)$$

8. A Gaussian pulse is propagating in the $+x$-direction and $t = t_0$ the displacement is given by

$$y(x, t = t_0) = a \exp\left[-\frac{(x-b)^2}{\sigma^2}\right]$$

Find $y(x, t)$.

9. A sonometer wire is stretched with a tension of 1 N. Calculate the velocity of transverse waves if $\rho = 0.2$ g/cm.

10. The displacement associated with a three-dimensional wave is given by

$$\psi(x, y, z, t) = a \cos\left[\frac{\sqrt{3}}{2} kx + \tfrac{1}{2} ky - \omega t\right]$$

Show that the wave propagates along a direction making an angle 30° with the x-axis.

11. Obtain the unit vector along the direction of propagation for a wave, the displacement of which is given by

$$\psi(x, y, z, t) = a \cos[2x + 3y + 4z - 5t]$$

where x, y and z are measured in centimeters and t in seconds. What will be the wavelength and the frequency of the wave?

$$\left[\text{Ans: } \frac{2}{\sqrt{29}}\hat{x} + \frac{3}{\sqrt{29}}\hat{y} + \frac{4}{\sqrt{29}}\hat{z}\right]$$

12. At $t = 0$, a wave packet is described by the following displacement

$$\psi(x, t = 0) = a \exp\left(-\frac{x^2}{\sigma^2}\right) \cos kx$$

Express the above displacement as a superposition of plane waves and interpret it physically if $k\sigma \gg 1$.

13. Calculate the velocity of longitudinal elastic waves in aluminium ($Y = 0.70 \times 10^{11}$ N/cm^2, $\rho = 2.7$ g/cm^2). Compare this value of velocity with the speed of sound in air.

14. Show that in a perfect gas, the velocity of sound increases as the square root of the absolute temperature.

REFERENCES AND SUGGESTED READING

BRADDICK, H.J.J. (1965), *Vibrations, Waves and Diffractions*, McGraw-Hill Publishing Co., London (England).

CRAWFORD, F.S. (1968), *Waves and Oscillations, Berkeley Physics Course*, Vol. III, McGraw-Hill Book Co., New York (USA).

COULSON, C.A. (1955), *Waves*, Seventh edition, Oliver & Boyd Ltd., Edinburgh (England).

ELMORE, W.C. and HEALD, M.A. (1969), *Physics of Waves*, McGraw-Hill Publishing Co., Maidenhead (England).

JOOS, G. (1955), *Theoretical Physics* (translated by I.M. Freeman), Blackie & Son Ltd., London (England).

PAIN, H.J. (1968), *The Physics of Vibrations and Waves*, John Wiley & Sons, London (England).

Physical Science Study Committee (1967), *Physics*, D.C. Heath & Co., Boston, Mass. (USA).

SLATER, J.C. and FRANK, N.H. (1969), *Electromagnetism*, Dover Publications, New York (USA).

WALDRON, R.A. (1964), *Waves and Oscillations*, Van Nostrand Publishing Co., New York (USA).

9 Huygens' Principle and Its Applications

And could such spacious virtue find a grave,
Beneath the imposthumed bubble of a wave!

LAMB

9.1. Introduction

The wave theory of light was first put forward by Christian Huygens in 1678. During that period, everyone believed in Newton's corpuscular theory, which had satisfactorily explained the phenomena of reflection, refraction, the rectilinear propagation of light and the fact that light could propagate through vacuum. So empowering was Newton's authority that the scientists around Newton believed in the corpuscular theory much more than Newton himself; as such, when Huygens put forward his wave-theory, no one really believed him. On the basis of his wave theory, Huygens explained satisfactorily the phenomenon of reflection, refraction and total internal reflection and also provided a simple explanation of the then recently discovered birefringence (see Ch. 18). As we will see later, Huygens' theory predicted that the velocity of light in a medium (like water shall be less than the velocity of light in free space, which is just the converse of the prediction made from Newton's corpuscular theory (see Sec. 1.2).

The wave character of light was not really accepted until the interference experiments of Young and Fresnel (in the early part of the nineteenth century) which could only be explained on the basis of a wave theory. At a later date, the data on the speed of light through transparent media were also available which was consistent with the results obtained by using the wave theory. It should be pointed out that Huygens did not know, whether the light waves were longitudinal or transverse and also how they propagate through vacuum. It was only in the later part of the nineteenth century, when Maxwell propounded his famous electromagnetic theory, could the nature of light waves be understood properly.

9.2. Huygens' Theory

Huygens' theory is essentially based on a geometrical construction which allows us to determine the shape of the wave front at any time, if the shape of the wavefront at an earlier time is known. A wavefront is the locus of the points which are in the same phase; for example, if we drop a small stone in a calm pool of water, circular ripples spread out from the point of impact, each point on the circumference of the circle (whose centre is at the point of impact) oscillates with the same amplitude and same phase and thus we have a circular wavefront. On the other hand, if we have a point source emanating waves in an uniform isotropic medium, the locus of points which have the same amplitude and are in the same phase are spheres. In this case we have spherical wavefronts as shown in Fig. 9.1(a). At large distances from the source, a small portion of the sphere can be considered as a plane and we have, what is known as a plane wave [see Fig. 9.1(b)].

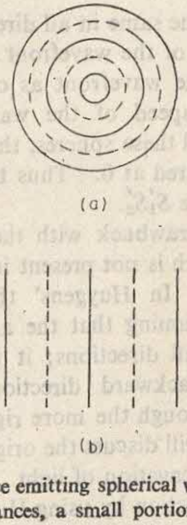


Fig. 9.1 (a) A point source emitting spherical waves.
(b) At large distances, a small portion of the spherical wavefront can be approximated to a plane wavefront thus resulting in plane waves.

Now, according to Huygens' principle, each point of a wavefront is a source of secondary disturbance and the wavelets emanating from these points spread out in all directions with the speed of the wave. The envelope of these wavelets give the shape of the new wavefront. In Fig. 9.2 S_1 and S_2 represent the shape of the wavefront (emanating from the point O) at a particular time which we denote as $t=0$. The medium is assumed to be homogeneous and isotropic, i.e. the medium is characterized by the same property at all points and the speed of

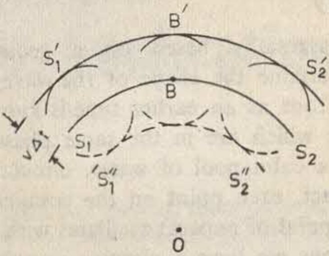


Fig. 9.2 Huygens' construction for the determination of the shape of the wavefront, given the shape of the wavefront at an earlier time. S_1S_2 is a spherical wavefront centered at O at any time, say $t=0$. $S_1'S_2'$ corresponds to the state of the wave front at a time Δt, which is again spherical centered at O. The dashed curve represents the backwave.

propagation of the wave is the same in all directions. Let us suppose we want to determine the shape of the wavefront after a time interval of Δt. Then, with each point on the wavefront as centre, we draw spheres of radius vΔt, where v is the speed of the wave in that medium. If we draw a common tangent to all these spheres, then we obtain the envelope which is again a sphere centred at 0. Thus the shape of the wavefront at a later time Δt is the sphere $S_1'S_2'$.

There is, however, one drawback with the above model, because we also obtain a backwave which is not present in practice. This backwave is shown as $S_1''S_2''$ in Fig. 9.2. In Huygens' theory, the presence of the backwave is avoided by assuming that the amplitude of the secondary wavelets is not uniform in all directions; it is maximum in the forward direction and zero in the backward direction.* The absence of the backwave is really justified through the more rigorous wave theory.

In the next section we will discuss the original argument of Huygens to explain the rectilinear propagation of light. In Sec. 9.4 we will derive the laws of refraction and reflection by using Huygens' principle. Finally, in Sec. 9.5 we will show how Huygens' principle can be used in inhomogeneous media.

*Indeed it can be shown from rigorous diffraction theory that one does obtain (under certain approximations) an obliquity factor, which is of the form $\frac{1}{2}(1+\cos\theta)$ where θ is the angle between the normal to the wavefront and the direction under consideration. Clearly when $\theta=0$, the obliquity factor is 1 (thereby giving rise to maximum amplitude in the forward direction) and when $\theta=\pi$, the obliquity factor is zero (thereby giving rise to zero amplitude in the backward direction).

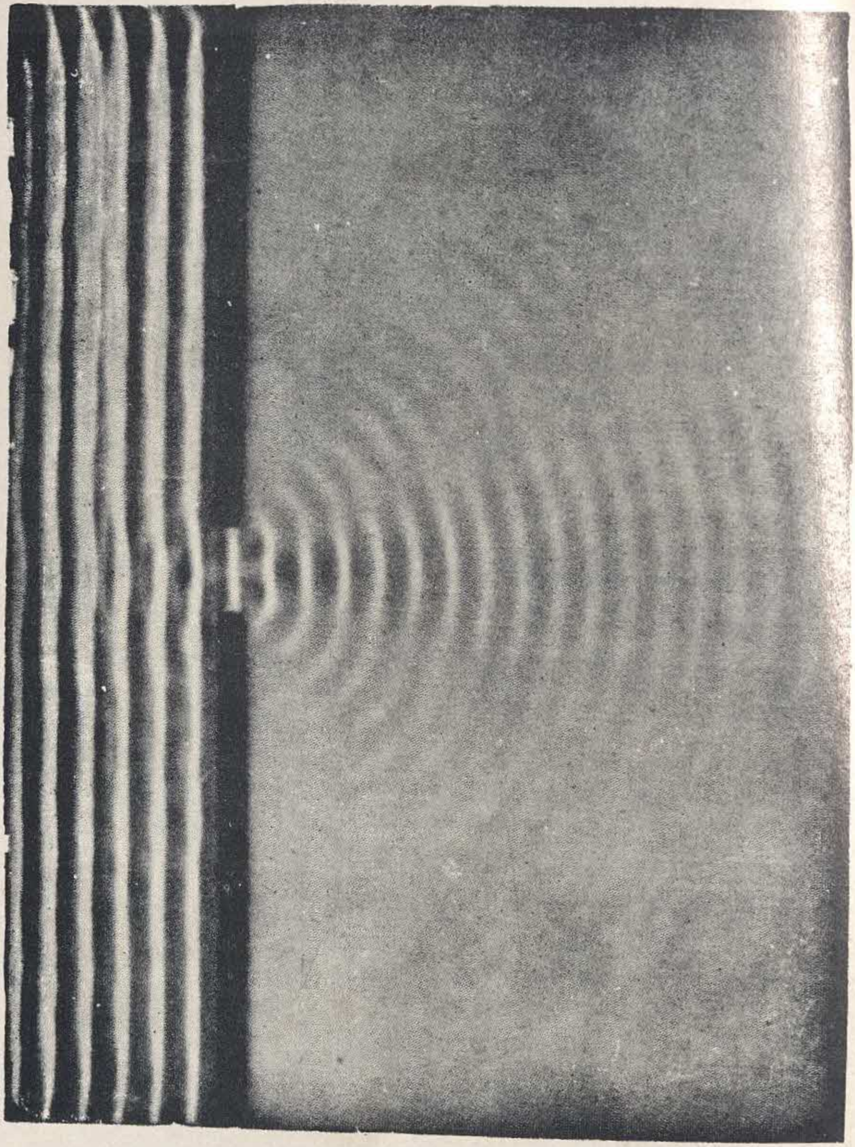


Plate 4 Diffraction of straight water waves when it passes through an opening. (After PSSC, *Physics*, D.C. Heath & Co., Boston, Mass., 1965; used with permission.)

9.3. Rectilinear Propagation

Let us consider spherical waves emanating from the point source *O* and striking the obstacle *A* (see Fig. 9.3). According to the rectilinear propagation of light (which is also predicted by corpuscular theory) one should obtain a shadow in the region *PQ* of the screen. As we will see in a later chapter, this is not rigorously true and one does obtain a finite intensity in the region of the geometrical shadow. However, at the time of Huygens, light was known to travel in straight lines and Huygens explained this by assuming that the secondary wavelets do not have any amplitude at any point not enveloped by the wavefront. Thus, referring back to Fig. 9.2, the secondary wavelets emanating from a typical point *B* will give rise to a finite amplitude at *B'* only and not at any other point.

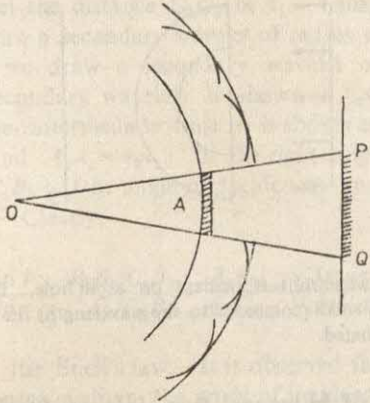


Fig. 9.3 Rectilinear propagation of light. *O* is a point source emitting spherical waves and *A* is an obstacle which forms a shadow in the region *PQ* of the screen.

The above explanation of the rectilinear propagation of light is indeed unsatisfactory and is incorrect. Further, as pointed out earlier, one does observe a finite intensity of light in the geometrical shadow. A satisfactory explanation was put forward by Fresnel, who postulated that the secondary wavelets mutually interfere. The Huygens' principle along with the fact that the secondary wavelets mutually interfere, is known as the Huygens-Fresnel principle. It may be mentioned that if a plane wave is allowed to fall on a tiny hole,* then the hole approximately acts as a point source and spherical waves emanate from it (see Fig. 9.4 and Plate 4). This fact is in direct contradiction to the original proposition of Huygens** according to which the secondary wavelets do not have any amplitude at any point not enveloped by the wavefront; however, as we will see in the chapter on diffraction, it can be explained satisfactorily on the basis of Huygens-Fresnel principle.

---

*By a tiny hole we imply that the diameter (of the hole) should be of the order of 0.1 mm or less.

**Use of the Huygens' principle in determining the shape of the wavefront in anisotropic media will be discussed in chapter 18.

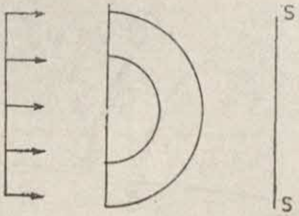


Fig. 9.4 A plane wavefront is incident on a pin hole. If the diameter of the pinhole is small (compared to the wavelength) the entire screen SS' will be illuminated.

9.4. Application of Huygens' Principle to Study Refraction and Reflection

(i) Refraction of a Plane Wave at a Plane Interface

We will first derive the laws of refraction. Let S_1S_2 be a surface separating two media with different speeds of propagation of light v_1 and v_2 as shown in Fig. 9.5. Let A_1B_1 be a plane wavefront incident on the surface at an angle i; A_1B_1 represents the position of the wavefront at an instant $t=0$. Let τ be the time taken for the wavefront to travel the distance B_1B_3. Then $B_1B_3=v_1\tau$. In the same time the light would have travelled a distance $A_1A_3=v_2\tau$ in the second medium. (Note that the lines A_1A_3, B_1B_3 etc. are always normal to the wavefront; these represent rays

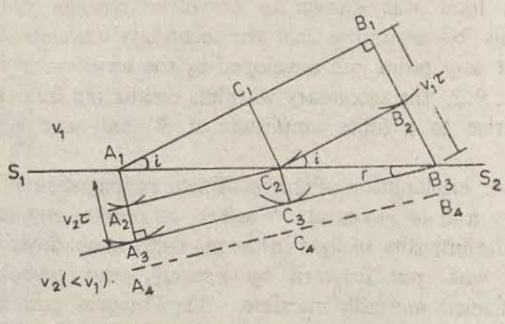


Fig. 9.5 Refraction of a plane wavefront A_1B_1 by a plane interface S_1S_2 separating two media with different velocities of propagation of light v_1 and v_2 ($<v_1$); i and r are the angles of incidence and refraction respectively. $A_2C_2B_2$ corresponds to the shape of the wavefront at an intermediate time τ_1. Notice that $r<i$.

in isotropic media (see Chapter 2). It can easily be seen that the incident and refracted rays make angles i and r with the normal. In order to determine the shape of the wavefront at the instant $t=\tau$ we consider an arbitrary point C_1 on the wavefront. Let the time taken for the disturbance to travel the distance C_1C_2 be τ_1. Thus $C_1C_2=v_1\tau_1$. From the point C_2 we draw a secondary wavelet of radius $v_2\ (\tau-\tau_1)$. Similarly from the point A_1 we draw a secondary wavelet of radius $v_2\tau$. The envelope of these secondary wavelets is shown as $A_3C_3B_3$. The shape of the wavefront at the intermediate time τ_1 is shown as $A_2C_2B_2$ and clearly $B_1B_2=C_1C_2=v_1\tau_1$ and $A_1A_2=v_2\tau_1$. In the right angled triangles $B_2C_2B_3$ and $C_3C_2B_3$, $\angle B_2C_2B_3=i$ (the angle of incidence) and $\angle C_2B_3C_3=r$ (the angle of refraction). Clearly,

$$\frac{\sin i}{\sin r}=\frac{B_2B_3/C_2B_3}{C_2C_3/C_2B_3}=\frac{B_2B_3}{C_2C_3}=\frac{v_1(\tau-\tau_1)}{v_2(\tau-\tau_1)}=\frac{v_1}{v_2} \tag{9.4-1}$$

which is known as the Snell's law. It is observed that when light travels from a rarer to a denser medium, the angle of incidence is greater than the angle of refraction and consequently

$$\frac{\sin i}{\sin r}>1$$

which implies $v_1>v_2$; thus, Huygens' theory predicts that the speed of light in a rarer medium is greater than the speed of light in a denser medium. This prediction is contradictory to that made by Newton's corpuscular theory (see Sec. 1.2) and as later experiments showed, the prediction of the wave theory was indeed correct.

If c represents the speed of light in free space then the ratio $\frac{c}{v}$ (where v represents the speed of light in the particular medium) is called the refractive index n of the medium. Thus if $n_1\left(=\frac{c}{v_1}\right)$ and $n_2\left(=\frac{c}{v_2}\right)$ are the refractive indices of the two media, then Snell's law can also be written as

$$n_1 \sin i = n_2 \sin r \tag{9.4-2}$$

Let $A_1C_1B_1$, $A_2C_2B_2$, $A_3C_3B_3$ and $A_4C_4B_4$ denote the successive positions of crests. If λ_1 and λ_2 denote the wavelength of light in medium 1 and medium 2 respectively then, the distance $B_1B_2(=B_2B_3=C_1C_2)$ will be equal to λ_1 and the distance $A_1A_2(=A_2A_3=C_2C_3)$ will be equal to λ_2. From Fig. 9.5 it is obvious that

$$\frac{\lambda_1}{\lambda_2}=\frac{\sin i}{\sin r}=\frac{v_1}{v_2} \tag{9.4-3}$$

or

$$v_1/\lambda_1=v_2/\lambda_2 \tag{9.4-4}$$

Thus, when a wave gets refracted into a denser medium ($v_1 > v_2$) the wavelength and the speed of propagation decrease but the frequency ($=v/\lambda$) remains the same; when refracted into a rarer medium the wavelength and the speed of propagation will increase. In Table 9.1 we have given the indices of refraction of several materials with respect to vacuum. In Table 9.2, the wavelength dependence of the refractive index for crown

TABLE 9.1

REFRACTIVE INDICES OF VARIOUS MATERIALS RELATIVE TO VACUUM

(For light of wave length $\lambda = 5.890 \times 10^{-5}$ cm)*

Material	n	Material	n
Vacuum	1.0000	Quartz (fused)	1.46
Air	1.0003	Rock salt	1.54
Water	1.33	Glass (ordinary crown)	1.52
Quartz (crystalline)	1.54	Glass (dense flint)	1.66

*After Arons (1963).

TABLE 9.2*

REFRACTIVE INDICES OF TELESCOPE CROWN GLASS AND VITREOUS QUARTZ FOR VARIOUS WAVELENGTHS

Wavelength	Telescope Crown	Vitreous Quartz
1. 6.562816×10^{-5} cm	1.52441	1.45640
2. 5.889953×10^{-5} cm	1.52704	1.45845
3. 4.861327×10^{-5} cm	1.53303	1.46318

NOTE: The wavelengths corresponding to serial numbers 1, 2 and 3 correspond roughly to the red, yellow and blue colours. The table shows the accuracy with which the wavelengths and refractive indices can be measured.

*Adapted from F. A. Jenkins and H. E. White, 3rd Edn., *Fundamentals of Optics*, McGraw-Hill (1957) p. 465.

glass and vitreous quartz are given. The three wavelengths correspond roughly to the red, yellow and blue colours. Notice the accuracy with which the wavelength and the refractive index can be measured.

(ii) Total Internal Reflection

In Fig. 9.5 the angle of incidence has been shown to be greater than the angle of refraction. This corresponds to the case when $v_2 < v_1$, i.e., the light wave is incident on a denser medium. However, if the second medium is a rarer medium (i.e. $v_1 < v_2$) then the angle of refraction will be greater than the angle of incidence, and a typical refracted wavefront would be of the form as shown in Fig. 9.6 where $B_1B_2 = v_1\tau$ and $A_1A_2 = v_2\tau$. Clearly, if the angle of incidence is such that $v_2\tau$ is greater than A_1B_2, then the refracted wavefront will be absent and we will have, what is known as, total internal reflection. The critical angle will correspond to

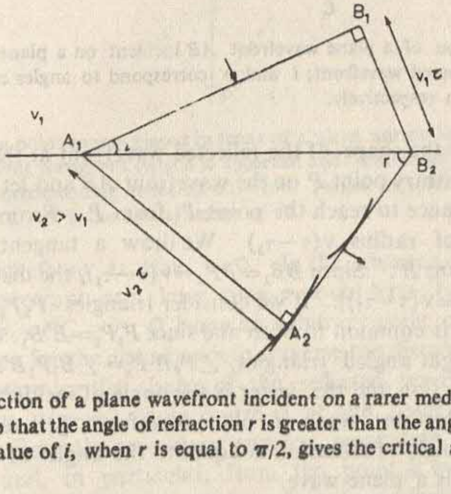


Fig. 9.6 Refraction of a plane wavefront incident on a rarer medium (i.e. $v_2 > v_1$). Notice that the angle of refraction r is greater than the angle of incidence i. The value of i, when r is equal to $\pi/2$, gives the critical angle.

$$A_1B_2 = v_2\tau$$

Thus

$$\sin i_c = \frac{B_1B_2}{A_1B_2} = \frac{v_1}{v_2} = n_{12}, \tag{9.4-5}$$

where i_c denotes the critical angle and n_{12} represents the refractive index of the second medium with respect to the first. For all angles of incidence greater than i_c, we will have total internal reflection.

(iii) Reflection of a Plane Wave by a Plane Surface

Let us consider a plane wave AB incident at an angle i on a plane mirror as shown in Fig. 9.7. We consider the reflection of the plane wave and try to obtain the shape of the reflected wavefront. Let the position of the wavefront at $t=0$ be AB. If the mirror was not present, then at a later time τ the position of the wavefront would have been CB', where $BB'=PP'=AC=v\tau$ and v is the speed of propagation of the wave. In

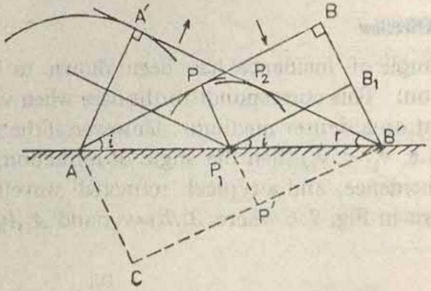


Fig. 9.7 Reflection of a plane wavefront AB incident on a plane mirror. $A'B'$ is the reflected wavefront; i and r correspond to angles of incidence and reflection respectively.

order to determine the shape of the reflected wavefront at the instant $t=\tau$, we consider an arbitrary point P on the wavefront AB and let τ_1 be the time taken by a disturbance to reach the point P_1 from P. From the point P_1, we draw a sphere of radius $v(\tau-\tau_1)$. We draw a tangent plane on this sphere from the point B'. Since $BB_1=PP_1=v(\tau-\tau_1)$, the distance B_1B' will be equal to $P_1P_2[=v(\tau-\tau_1)]$. If we consider triangles P_2P_1B' and B_1P_1B' then the side P_1B' is common to both and since $P_1P_2=B'B_1$, and since both the triangles are right angled triangles, $\angle P_2B'P_1=\angle B_1P_1B'$. The former is the angle of reflection and the latter is the angle of incidence. Thus, we have the law of reflection; when a plane wavefront gets reflected from a plane mirror, the angle of reflection is equal to the angle of incidence and the reflected wave is a plane wave.

(iv) Diffuse Reflection

In the above we have considered the reflection of light from a smooth surface. This is known as specular reflection. If the surface is irregular (as shown in Fig. 9.8) we have, what is known as, diffuse reflection. The

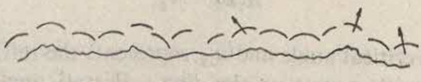


Fig. 9.8 Diffuse reflection of a plane wavefront from a rough surface. It is evident that one does not have a well-defined reflected beam.

secondary wavelets emanating from the irregular surface travel in many directions and we do not have a well defined reflected wave. Indeed, it can be shown that if the irregularity in the surface is considerably greater than the wavelength, we will have diffuse reflection.

(v) Reflection of Light from a Point Source Near a Mirror

Let us consider spherical waves (emanating from a point source P) incident on a plane mirror MM', as shown in Fig. 9.9. Let ABC denote the shape of the wavefront at time $t=0$. In the absence of the mirror, the shape of the wavefront at a later time τ would have been $A_1B_1C_1$ where $AA_1=QQ_1=BB_1=CC_1=v\tau$, Q being an arbitrary point on the wavefront. If the time taken for the disturbance to traverse the distance QQ' be τ_1 then, in order to determine the shape of the reflected wavefront, we draw a sphere of radius $v(\tau-\tau_1)$ whose centre is at the point Q'. In a similar manner we can draw the secondary wavelets emanating from other points on the mirror and, in particular, from the point B we have to draw a sphere of radius $v\tau$. The shape of the reflected wavefront is obtained by drawing a common tangent plane to all these spheres, which is shown as $A_1B_1'C_1$ in the figure. It can immediately be seen that $A_1B_1'C_1$ will have an exactly similar shape as $A_1B_1C_1$ except that $A_1B_1'C_1$ will have its centre of curvature at the point P' where $PB=BP'$. Thus the reflected waves will appear to emanate from the point P' which will be the virtual image of the point P.

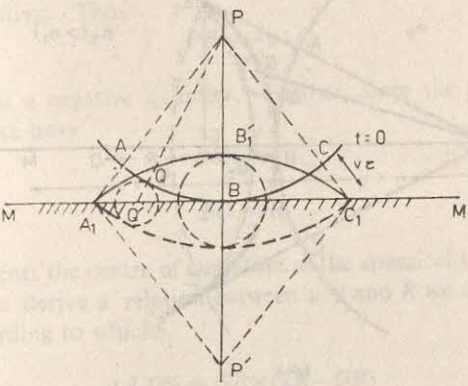


Fig. 9.9 P is a point source placed in front of a plane mirror MM'. ABC is the incident wavefront (which is spherical and centred at P) and $A_1B_1'C_1$ is the corresponding reflected wavefront (which is spherical and centred at P'). P' is the virtual image of P.

(vi) Refraction of a Spherical Wave by a Spherical Surface

Let us consider spherical waves (emanating from the point P) incident on the curved spherical surface SBS'. Let the shape of the wavefront at the time $t=0$ be ABC [see Fig. 9.10 (a)]. Let the refractive indices on the left and on the right of the spherical surface be n_1 and n_2 respectively. In the absence of the spherical surface, the shape of the wavefront at a later time τ would have been $A_1B_1C_1$ where $AA_1=BB_1=CC_1=v_1\tau$. We consider an arbitrary point Q on the wavefront ABC and let τ_1 be the time taken for the disturbance to reach the point Q' (on the surface of the spherical wave); thus $QQ'=v_1\tau_1$. In order to determine the shape of the refracted wavefront

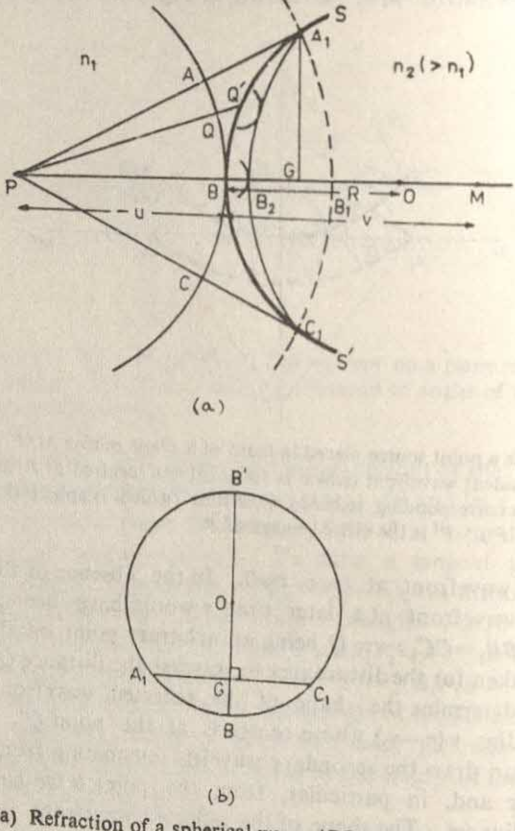


Fig. 9.10 (a) Refraction of a spherical wave *ABC* (emanating from the point source *P*) by a convex spherical surface *SBS'* separating media of refractive indices n_1 and n_2 $(>n_1)$. $A_1B_2C_1$ is the refracted wavefront, which is approximately spherical and whose centre of curvature is at *M*. Thus *M* is the real image of *P*. *O* is the centre of curvature of *SS'*.
(b) The diameter *B'OB* intersects the chord A_1GC_1 normally.

at a later time τ, we draw a sphere of radius $v_2(\tau-\tau_1)$ from the point Q'. We may draw similar spheres from other points on the spherical surface; in particular, the radius of the spherical wavefront from the point *B* which is equal to BB_2, will be $v_2\tau$. The envelope of these spherical wavelets is shown as $A_1B_2C_1$ which, in general, will not be a sphere.* However, a small portion of any curved surface can be considered as a sphere and in this approximation we may consider $A_1B_2C_1$ to be a sphere whose centre of curvature is

*The fact that the refracted wavefront is not, in general, a sphere leads to, what are known as, aberrations.

at the point M. The spherical wavefront will, therefore, converge towards the point M and hence the point M represents the real image of the point P.

We adopt a sign convention in which all distances, measured to the left of the point B are negative and all distances measured to the right of the point B are positive. Thus

$$PB = -u$$

where u itself is a negative quantity. Further, since the point M lies on the right of B, we have

$$BM = v$$

and similarly,

$$BO = R$$

where O represents the centre of curvature of the spherical surface.

In order to derive a relation between u, v and R we use a theorem in geometry, according to which,

$$(A_1G)^2 = GB \times (2R - GB) \tag{9.4-6}$$

where G is the foot of the perpendicular on the axis PM [See Fig. 9.10(b)]. In Fig. 9.10(b) the diameter $B'OB$ intersects the chord A_1GC_1 normally. If $GB \ll R$, then

$$(A_1G)^2 \simeq 2R(GB)$$

Consider the spherical surface SBS' [Fig. 9.10(a)] whose radius is R. Clearly,

$$\begin{aligned}(A_1G)^2 &= (2R - GB)\,GB \\ &\simeq 2R(GB)\end{aligned} \tag{9.4-7}$$

where we have assumed $GB \ll R$. Similarly by considering the spherical surface $A_1B_2C_1$ (whose centre is at the point M) we obtain

$$(A_1G)^2 \simeq 2v(GB_2) \tag{9.4-8}$$

where $v = BM \simeq B_2M$. In a similar manner,

$$(A_1G)^2 \simeq 2(-u)\,GB_1 \tag{9.4-9}$$

Since u is a negative quantity, $(A_1G)^2$ is positive. Now

$$BB_1 = v_1\tau \quad \text{and} \quad BB_2 = v_2\tau$$

Therefore,

$$\frac{BB_1}{BB_2} = \frac{v_1}{v_2} = \frac{n_2}{n_1}$$

or

$$n_1 BB_1 = n_2 BB_2$$

or

$$n_1(BG + GB_1) = n_2(BG - GB_2)$$

or

$$n_1\left[\frac{(A_1G)^2}{2R}-\frac{(A_1G)^2}{2u}\right]=n_2\left[\frac{(A_1G)^2}{2R}-\frac{(A_1G)^2}{2v}\right]$$

where we have used Eqs. (9.4-7), (9.4-8) and (9.4-9). Thus

$$\frac{n_2}{v}-\frac{n_1}{u}=\frac{n_2-n_1}{R} \tag{9.4-10}$$

which may be rewritten in the form

$$\frac{n_2}{v}=\frac{n_1}{u}+\frac{n_2-n_1}{R} \tag{9.4-11}$$

Thus, if

$$\frac{n_1}{|u|}>\frac{n_2-n_1}{R}$$

or

$$|u|<\frac{Rn_1}{n_2-n_1}$$

we will obtain a virtual image. (We are of course assuming that the second medium is a denser medium, i.e. $n_2>n_1$; if $n_2<n_1$, we will always have a virtual image.)

It may be pointed out that a converging spherical wavefront will propagate in a manner shown in Fig. 9.11. Beyond the focal point it will start diverging as shown in the figure.*

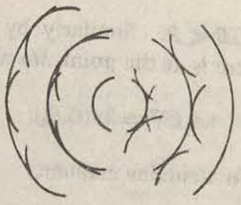


Fig. 9.11 Propagation of a converging spherical wave using Huygens' principle.

In a similar manner we can consider the refraction of a spherical wave from a surface SBS' shown in Fig. 9.12 ($n_2>n_1$). Here the centre of curvature will also lie on the left of the point B and both u and R will be negative quantities. Thus no matter what the values of u and R may be, v will be negative and we will obtain a virtual image.

---

*Very close to the focal point, one has to use a more rigorous wave theory and the shape of the wavefront is very much different from spherical (see Born and Wolf, 1975). However, much beyond the focal point the wavefronts again become spherical.

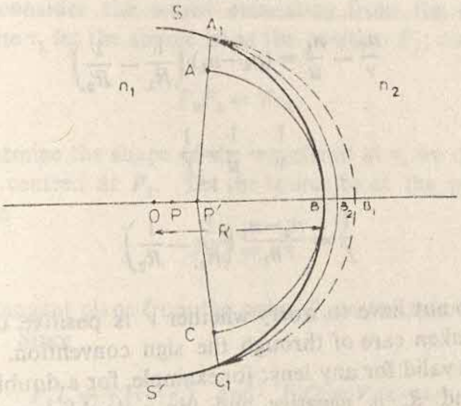


Fig. 9.12 Refraction of a spherical wave by a concave surface separating media of refractive indices n_1 and n_2 ($> n_1$). P' is the virtual image of P.

Using Eq. (9.4-10) we can easily derive the thin lens formula. We assume a thin lens made of a material of refractive index n_2 to be placed in a medium of refractive index n_1 (See Fig. 9.13). Let the radii of curvatures

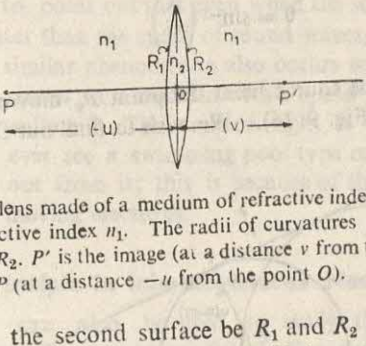


Fig. 9.13 A thin lens made of a medium of refractive index n_2 placed in a medium of refractive index n_1. The radii of curvatures of the two surfaces are R_1 and R_2. P' is the image (at a distance v from the point O) of the point object P (at a distance $-u$ from the point O).

of the first and the second surface be R_1 and R_2 respectively. Let v' be the distance of the image of the object P if the second surface were not present. Then

$$\frac{n_2}{v'} - \frac{n_1}{u} = \frac{n_2 - n_1}{R_1} \tag{9.4-12}$$

(Since the lens is assumed to be thin, all the distances are measured from the point O.) This image now acts as an object to the spherical surface R_2 on the left of which is the medium of refractive index n_2 and on the right of which is the medium of refractive index n_1. Thus, if v is the distance of the final image point from O, then

$$\frac{n_1}{v} - \frac{n_2}{v'} = \frac{n_1 - n_2}{R} \tag{9.4-13}$$

Adding Eqs. (9.4-12) and (9.4-13) we obtain

$$\frac{n_1}{v} - \frac{n_1}{u} = (n_2 - n_1)\left(\frac{1}{R_1} - \frac{1}{R_2}\right) \tag{9.4-14}$$

or,

$$\frac{1}{v} - \frac{1}{u} = \frac{1}{f} \tag{9.4-15}$$

where

$$\frac{1}{f} = \frac{n_2 - n_1}{n_1}\left(\frac{1}{R_1} - \frac{1}{R_2}\right) \tag{9.4-16}$$

Notice that we do not have to worry whether v' is positive or negative; it is automatically taken care of through the sign convention. Further, the relation derived is valid for any lens; for example, for a double convex lens, R_1 is positive and R_2 is negative and for a double concave lens, R_1 is negative and R_2 is positive. Similarly it follows for other types of lenses (see Fig. 3.6).

Example 9.1 Consider a vibrating source moving through a medium with a speed V. Let the speed of propagation of the wave in the medium be v. Show that if $V >$ v then a conical wavefront is set up whose half-angle is given by

$$\theta = \sin^{-1}\left(\frac{\text{v}}{V}\right) \tag{9.4-17}$$

Solution. Let at $t=0$, the source be at the point P_0 moving with a speed V in the x-direction (see Fig. 9.14). We wish to find out the wavefront at

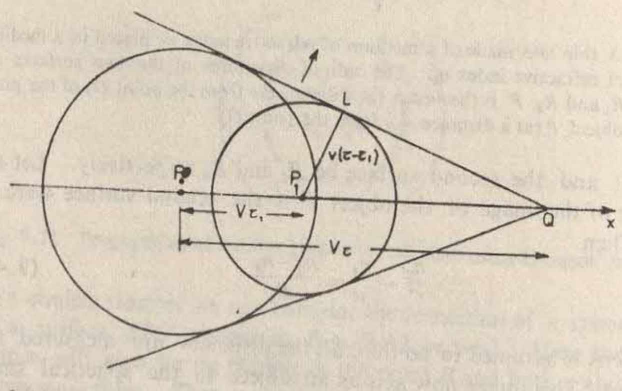


Fig. 9.14 Production of a shock wavefront by a vibrating particle P_0 moving with a speed V, in a medium in which the velocity of propagation of the wave is v ($< V$).

a later time τ. The disturbance emanating from the point P_0 traverses a distance vτ in time τ. Thus from the point P_0 we draw a sphere of radius vτ. We next consider the waves emanating from the source at a time τ_1 ($<\tau$). At time τ_1 let the source be at the position P_1; consequently

$$P_0P_1 = V\tau_1$$

In order to determine the shape of the wavefront at τ, we draw a sphere of radius v $(\tau-\tau_1)$ centred at P_1. Let the source be at the position Q at the instant τ. Then

$$P_0Q = V\tau$$

We draw a tangent plane from the point Q, on the sphere whose origin is the point P_1. Since

$$P_1L = \text{v}\,(\tau-\tau_1) \quad \text{and} \quad P_1Q = V(\tau-\tau_1)$$

$$\sin\theta = \frac{P_1L}{P_1Q} = \frac{\text{v}}{V} \text{(independent of } \tau_1)$$

Since θ is independent of τ_1, all the spheres drawn from any point on the line P_0Q will have a common tangent plane. This plane, is known as the shock wavefront and propagates with a speed v.

It is interesting to point out that even when the source is not vibrating, if its speed is greater than the speed of sound waves, a shock wavefront is always set up. A similar phenomenon also occurs when a charged particle (like an electron) moves in a medium with a speed greater than the speed of light in that medium.* The emitted light is known as Cerenkov radiation. If you ever see a swimming pool type reactor, you will find a blue glow coming out from it; this is because of the Cerenkov radiation emitted by the fast moving electrons.

9.5 Huygens' Principle in Inhomogeneous Media

Huygens' principle can also be used to study the propagation of a wavefront in an inhomogeneous medium. For definiteness, we consider a medium whose refractive index decreases continuously from a given axis, which we define as the z-axis; x-and y-axes being the transverse axes. A simple example is a Selfoc fibre,** whose refractive index variation is of the form

$$n(x, y) = n_0 - n_2(x^2 + y^2) \tag{9.5-1}$$

---

*This does not contradict the theory of relativity according to which no particle can have a speed greater than the speed of light in free space ($=3\times10^8$ m/sec). The speed of light in a medium will be equal to c/n, where n represents the refractive index. For example in water, the speed of light will be about $2{\cdot}25\times10^8$ m/sec and the speed of the electron could be greater than this value.

**See Example 2.4.

where n_0 is the refractive index on the z-axis. Let the plane wavefront be incident along the z-axis as shown in Fig. 9.15. Since the refractive index decreases as x and y increase, the speed of the secondary wavelets emanating from portions of the incident wavefront will increase as we move away from the axis. Let us try to determine the shape of the wavefront at a time Δt; given that the wavefront at $t=0$ is a plane wavefront A_1B_1 (see Fig. 9.15). We will have to draw spheres of radius $v(x, y)\,\Delta t$, centered at (x, y), where $v(x, y)$ is the velocity of the wave at the point (x, y), which increases as x and y increase. Thus the radii of the spheres increase as we move away from the axis and if we draw a common tangent to all these spheres then the resulting wavefront is shown in Fig. 9.15 as A_2B_2. It is at once evident that the wavefront which was initially plane has now become curved. If we again use the same procedure, then the shape of the wavefront at time $2\Delta t$ (say) is shown as A_3B_3. Thus it is evident that in the present case the wavefront is getting focussed. It should be borne in mind that since we are considering an inhomogeneous medium, the refractive index varies continuously with position. For the above construction to be valid, Δt should be small so that during this short interval the secondary wavelets may be assumed to be spherical; this treatment will be rigorously valid in the limit of $\Delta t \to 0$.

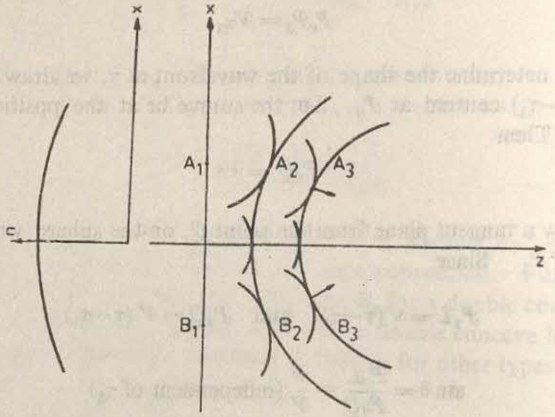


Fig. 9.15 The focusing of an incident plane wavefront in an inhomogeneous medium characterized by a refractive index variation given by Eq. (9.5-1).

PROBLEMS

1. Use Huygens' principle to study the reflection of a spherical wave (emanating from a point on the axis) at a concave mirror of radius of curvature R and obtain the mirror equation

$$\frac{1}{u}+\frac{1}{v}=\frac{2}{R}$$

2. Consider a plane wave incident obliquely on the face of a prism. Using Huygens' principle, construct the transmitted wavefront and show that the deviation produced by the prism is given by

$$\delta = i+t-A$$

where A is the angle of the prism, i and t are the angles of incidence and transmittance.

3. In an anisotropic medium (like calcite) the velocity of the extraordinary wave is different in different directions (see Sec. 2.4 and 18.5); indeed, one can show that the angular variation of velocity is given by the following equation

$$\frac{1}{v^2(\theta)}=\frac{1}{v_x^2\sin^2\theta}+\frac{1}{v_z^2\cos^2\theta}$$

where v_x and v_z are constants and the angle θ is measured with respect to a fixed direction which is known as the optic axis. Show that the wavefronts emanating from a point source will be ellipsoids of revolution.

4. A plane wave is incident normally on an anisotropic medium like calcite (see previous problem). Show that the refracted waves are also plane waves propagating in the same direction but the ray direction is, in general, not normal to the wavefronts. (The ray is along the line joining the source of secondary wavelet to the point of contact of the secondary wavelet with the envelope).

5. In the above problem discuss the formation of the refracted wave for a plane wave incident obliquely on an anisotropic media whose optic axis lies in the plane of incidence.

6. In a medium characterized by the following refractive index variation

$$n^2(x) = n_0^2+n_p^2\,[1-e^{-\alpha x}]$$

[see Eq. (2.3-12)]; discuss using Huygens' principle the bending of the beam around the point B in Fig. 2.9; this point is known as the turning point.

REFERENCES AND SUGGESTED READING

ARONS, A.B. (1965), *Development of Concepts in Physics*, Addison-Wesley Publishing Co., Reading, Mass. (USA).

BAKER, B.B. AND COPSON, E.J., (1969), *The Mathematical Theory of Huygens' Principle*, Oxford University Press, London (England).

BRADDICK, H.J.J. (1965), *Vibrations, Waves and Diffraction*, McGraw-Hill Publishing Co., London (England).

DEWITTE, A.J. (1959), Equivalence of Huygens' principle and Fermat's principle in ray geometry, *Amer. J. Phys.*, **27**, 293.

HUYGENS, C. (1962), *Treatise on Light*, Dover Publications, New York (USA).

PART III

Interference

10 Superposition of Waves

'If you are dropping pebbles into a pond and do not watch the spreading rings, your occupation should be considered as useless', said the fictional Russian philosopher, Kuzma Prutkoff. And, indeed we can learn much by observing these graceful circles spreading out from the punctured surface of calm water.

GAMOW AND CLEVELAND

10.1 Introduction

In this chapter we will discuss the applications of the principle of superposition of waves according to which the resultant displacement (at a particular point) produced by a number of waves is the vector sum of the displacements produced by each one of the disturbances. As a simple example, we consider a long stretched string AB (see Fig, 10.1). From the

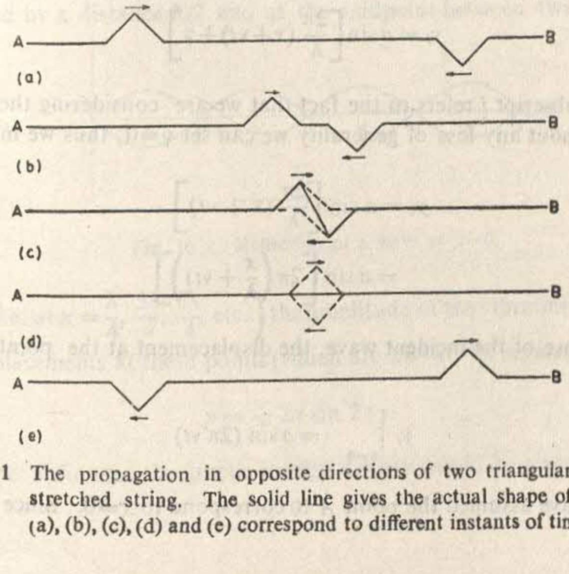


Fig. 10.1 The propagation in opposite directions of two triangular pulses in a stretched string. The solid line gives the actual shape of the string; (a), (b), (c), (d) and (e) correspond to different instants of time.

end A, a triangular pulse is generated which propagates to the right with a certain speed v. In the absence of any other disturbance, this pulse would have propagated in the $+x$-direction without any change in shape; we are, of course, neglecting any attenuation or distortion of the pulse. We next assume that from the end B an identical pulse is generated which starts moving to the left with the same speed v. (As has been shown in Sec. 8.6 the speed of the wave is determined by the ratio of the tension in the string to its mass per unit length). At $t=0$, the snapshot of the string is shown in Fig. 10.1 (a). At a little later time each pulse moves close to the other as shown in Fig. 10.1 (b); without any interference. Fig. 10.1 (c) represents a snapshot at an instant when the two pulses interfere; the dashed curves represent the profile of the string if each of the impulse was moving all by itself, whereas the solid curve shows the resultant displacement obtained by algebraic addition of each displacement. Shortly later [Fig. 10.1 (d)] the two pulses exactly overlap each other and the resultant displacement is zero everywhere (where has the energy gone?). At a much later time the impulses sort of cross each other [Fig. 10.1 (e)] and move as if nothing had happened. This is a characteristic feature of superposition of waves.

The phenomenon of interference contains no more physics than embodied in the above example. In the following sections we will consider some more examples.

10.2 Stationary Waves on a String

Consider a string which is fixed at the point A (see Fig. 10.2). A transverse sinusoidal wave is sent down the string along the $-x$-direction. The displacement at any point on the string due to this wave would be given by

$$y_i = a \sin\left[\frac{2\pi}{\lambda}(x+vt)+\varphi\right] \tag{10.2-1}$$

where the subscript i refers to the fact that we are considering the incident wave. Without any loss of generality we can set $\varphi=0$; thus we may write

$$\begin{aligned} y_i &= a \sin\left[\frac{2\pi}{\lambda}(x+vt)\right] \\ &= a \sin\left[2\pi\left(\frac{x}{\lambda}+\nu t)\right)\right] \end{aligned} \tag{10.2-2}$$

Thus, because of the incident wave, the displacement at the point A would have been

$$y_i\Big|_{x=0} = a \sin(2\pi \nu t) \tag{10.2-3}$$

where we have assumed the point A to correspond to $x=0$. Since the point

A is fixed, there must be a reflected wave such that the displacement due to this reflected wave (at the point A) is equal and opposite to y_i :

$$y_r \Big|_{x=0} = -a \sin (2\pi \nu t) \tag{10.2-4}$$

where the subscript r refers to the fact that we are considering the reflected wave. Since the reflected wave propagates in the $+x$-direction we must have

$$y_r = +a \sin 2\pi \left(\frac{x}{\lambda} - \nu t\right) \tag{10.2-5}$$

The resultant displacement would be given by

$$\begin{aligned} y = y_i + y_r &= a \left[\sin 2\pi \left(\frac{x}{\lambda} + \nu t\right) + \sin 2\pi \left(\frac{x}{\lambda} - \nu t\right) \right] \\ &= 2a \sin \frac{2\pi}{\lambda} x \sin 2\pi \nu t \end{aligned} \tag{10.2-6}$$

It should be seen that for values of x such that

$$\sin \frac{2\pi}{\lambda} x = 0 \tag{10.2-7}$$

the displacement y is zero *at all times*. Such points are known as *nodes*; the x-coordinates of the nodes are given by

$$x = 0, \frac{\lambda}{2}, \lambda, \frac{3\lambda}{2}, 2\lambda, \ldots \tag{10.2-8}$$

and are marked as points A, P, Q and R in Fig. 10.2. The nodes are separated by a distance $\lambda/2$ and at the midpoint between two consecutive

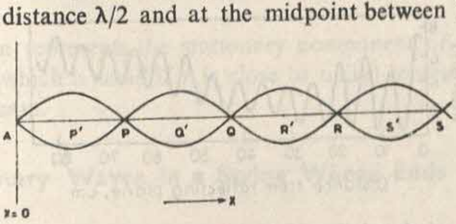


Fig. 10.2 Reflection of a wave at $x=0$.

nodes $\left(\text{i.e. at } x = \frac{\lambda}{4}, \frac{3\lambda}{4}, \frac{5\lambda}{4}, \text{etc.}\right)$ the amplitude of the vibration is maximum. The displacements at these points (which are known as *antinodes*) are given by

$$y = \pm 2a \sin 2\pi \nu t \tag{10.2-9}$$

At the antinodes the kinetic energy density would be given by (see Sec. 5.2) :

$$\begin{aligned}\text{Kinetic energy/unit length} &= \tfrac{1}{2}\rho(2a)^2\omega^2\cos^2\omega t\\ &= 2\rho a^2\omega^2\cos^2\omega t \qquad (10.2\text{-}10)\end{aligned}$$

where $\omega = 2\pi\nu$ is the angular frequency and ρ the mass per unit length of the string.

We can also carry out a similar experiment for electromagnetic waves (Fig. 10.3). In Fig. 10.3, *T* represents a transmitter of electromagnetic waves (the wavelength of which may be of the order of few cm); *R* represents a reflector which may be a highly polished metal surface and *D* represents the detector which can measure the variation of the intensity of the electromagnetic waves at different points. One may approximately assume plane waves to be incident on the reflector which get reflected; the incident and reflected waves interfere and produce nodes and anti-nodes. The result of a typical experiment is shown in Fig. 10.4. One can see the periodic variation of intensity. Two consecutive maxima are separated by about 5.8 cm; thus $\lambda \approx 11.6$ cm. The corresponding frequency ($\approx 2.6\times10^9$ sec^{-1}) can easily be generated in the laboratory. If the frequency is changed, one can observe the change in the distance between the antinodes. One should notice that the minima do not really correspond to zero intensity and the intensities at the maxima are not constant. This is because of the

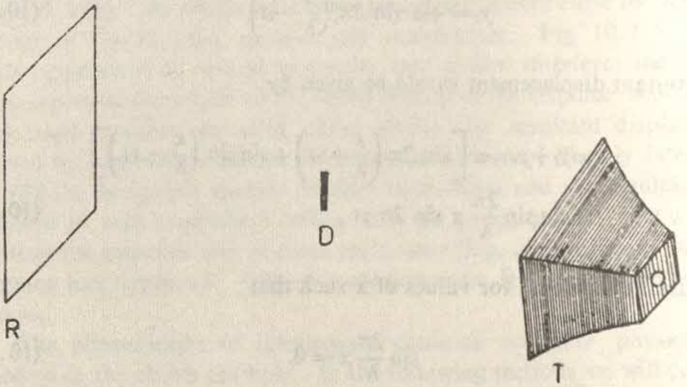


Fig. 10.3 An arrangement for studying standing electromagnetic waves.

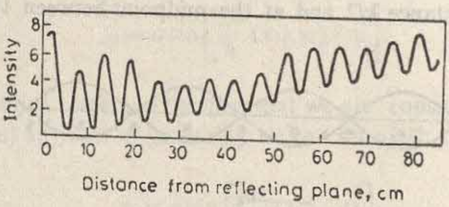


Fig. 10.4 A typical variation of the intensity between the reflector and the transmitter [adapted from Ditchburn (1963), p. 63].

fact that the incident wave is really not a plane wave* and that the reflection is not really perfect. In fact, one can introduce a coefficient of reflection (r) which is defined as the ratio of the energy of the reflected beam to the energy of the incident beam. Thus the ratio of the amplitudes would be $\sqrt{r}$ and if the incident wave is given by

$$E_{\text{incident}} = a \sin\left[2\pi\left(\frac{x}{\lambda}+\nu t\right)\right] \tag{10.2-11}$$

then the reflected field would be given by

$$E_{\text{reflected}} = a\sqrt{r} \sin\left[2\pi\left(\frac{x}{\lambda}-\nu t\right)\right] \tag{10.2-12}$$

where the plane $x = 0$ corresponds to the plane of the reflector. Here E represents the electric field associated with the electromagnetic wave. Thus the resultant field would be given by

$$\begin{aligned}
E_{\text{resultant}} &= E_{\text{incident}} + E_{\text{reflected}} \\
&= a \sin\left[2\pi\left(\frac{x}{\lambda}+\nu t\right)\right] + a\sqrt{r} \sin\left[2\pi\left(\frac{x}{\lambda}-\nu t\right)\right] \\
&= a\sqrt{r}\left\{\sin\left[2\pi\left(\frac{x}{\lambda}+\nu t\right)\right] + \sin\left[2\pi\left(\frac{x}{\lambda}-\nu t\right)\right]\right\} \\
&\qquad + a(1-\sqrt{r}) \sin\left[2\pi\left(\frac{x}{\lambda}+\nu t\right)\right] \\
&= 2a\sqrt{r} \sin\left(\frac{2\pi}{\lambda}x\right)\cos 2\pi\nu t \\
&\qquad + a(1-\sqrt{r}) \sin\left[2\pi\left(\frac{x}{\lambda}+\nu t\right)\right]
\end{aligned} \tag{10.2-13}$$

The first term represents the stationary component of the wave and the second term (which is small if r is close to unity) represents the progressive part of the beam.

10.3 Stationary Waves in a String Whose Ends Are Fixed

In Sec. 10.2, while discussing the stationary waves on a string we had assumed only one end of the string ($x = 0$) to be fixed; and the resultant displacement was shown to be given by [see Eq. (10.2-6)]:

$$y = 2a \sin\left(\frac{2\pi}{\lambda}x\right)\sin(2\pi\nu t) \tag{10.3-1}$$

---

*A plane wave is obtained by a point source at a very large distance from the point of observation (see Chapter 8).

If the other end of the string (say at $x=L$) is also fixed then we must have

$$2a \sin\left(\frac{2\pi}{\lambda} L\right) \sin(2\pi\nu t)=0 \qquad (10.3\text{-}2)$$

Eq. (10.3-2) is to be valid at all times, therefore.

$$\sin\left(\frac{2\pi}{\lambda} L\right)=0=\sin n\pi \qquad (10.3\text{-}3)$$

or

$$\lambda_n=\frac{2L}{n},\ n=1, 2, 3,\ldots. \qquad (10.3\text{-}4)$$

The corresponding frequencies are

$$\nu_n=\frac{v}{\lambda_n}=\frac{nv}{2L},\ n=1, 2, 3,\ldots. \qquad (10.3\text{-}5)$$

Thus, if a string (of length L) is clamped at both ends (as in a sonometer wire) then it can only vibrate with certain well defined wavelengths. When $\lambda=2L$ (i.e. $n=1$) the string is said to vibrate in its fundamental mode [Fig. 10.5(a)]. Similarly when $\lambda=2L/2$ and $2L/3$ the string is said to vibrate in its first and second harmonic. In general, if the string is plucked and then made to vibrate then the displacement would be given by

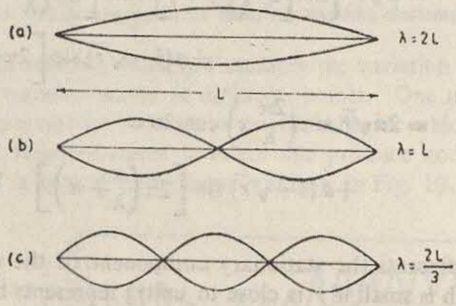


Fig. 10.5 Standing waves on a stretched string clamped at both ends.

$$y(x, t)=\sum_{n=1}^{\infty} a_n \sin\left(\frac{2\pi}{\lambda_n} x\right) \sin(2\pi\nu_n t+\varphi_n) \qquad (10.3\text{-}6)$$

where the constants a_n and φ_n are determined by the values of $y(x, t=0)$ and $\left.\frac{\partial y}{\partial t}\right|_{t=0}$; these are known as the initial conditions. A more detailed discussion on the vibration of stretched strings has been given in Chapter 7.

It should be seen that when a string is vibrating in a particular mode there is no net transference of energy although each element of the string is associated with a certain energy density (see Eq. 10.2-10). The energy density is maximum at the antinodes and minimum at nodes. The distances between two successive antinodes and successive nodes are $\lambda/2$.

10.4 Stationary Light Waves: Ives and Wiener's Experiments

It is difficult to carry out experiments in which one obtains stationary light waves. This is because of the fact that light wavelengths are extremely small ($\approx 5\times10^{-5}$cm). In the experimental arrangement of Ives, the emulsion side of a photographic plate was placed in contact with a film of mercury as shown in Fig. 10.6. A parallel beam of monochromatic light was allowed to fall normally on the glass plate. The beam was reflected on the mercury surface and the incident wave interfered with the reflected wave forming standing waves. A section of the photographic film was cut along a plane normal to the surface. The cut section was viewed under a microscope and bright and dark bands (separated by regular intervals) were observed. By measuring the distance between two consecutive dark bands (which is equal to $\lambda/2$) one can calculate the wavelength.

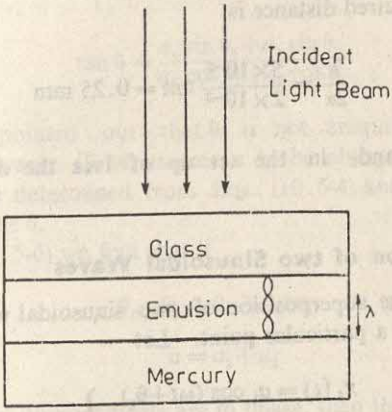


Fig. 10.6 The experimental arrangement of Ives for studying stationary light waves.

Because of the small wavelength of light, the distance between two consecutive dark (or bright) bands was extremely small and was, therefore, difficult to measure. Wiener overcame this difficulty by placing the photographic film at a small angle and thereby increasing considerably the distance between the dark (or bright) bands (Fig. 10.7).

Example 10.1 In a typical experimental arrangement of Wiener, the angle between the film and the mirror was about 10^{-3} radians. For

$\lambda \approx 5 \times 10^{-5}$ cm what would be the distance between two consecutive dark bands?

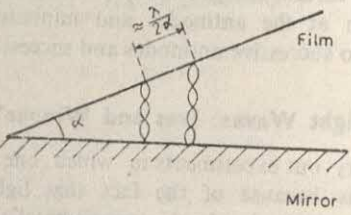


Fig. 10.7 The experimental arrangement of Wiener for studying stationary light waves.

Solution. The required distance is

$$\frac{\lambda}{2\alpha} = \frac{5 \times 10^{-5}}{2 \times 10^{-3}} \text{ cm} = 0.25 \text{ mm}$$

On the other hand, in the set up of Ives the distance would be 2.5×10^{-4} mm.

10.5 Superposition of two Sinusoidal Waves

Let us consider the superposition of two sinusoidal waves (having the same frequency*) at a particular point. Let

$$\left.\begin{aligned} x_1(t) &= a_1 \cos(\omega t + \theta_1) \\ \text{and} \qquad x_2(t) &= a_2 \cos(\omega t + \theta_2) \end{aligned}\right\} \qquad (10.5\text{-}1)$$

represent the displacements produced by each of the disturbances: we are assuming that the displacements are in the same direction,** however, they may have different amplitudes and different initial phases. Now, according to the superposition principle the resultant displacement $x(t)$ would be given by

$$\begin{aligned} x(t) &= x_1(t) + x_2(t) \\ &= a_1 \cos(\omega t + \theta_1) + a_2 \cos(\omega t + \theta_2) \end{aligned} \qquad (10.5\text{-}2)$$

*In Chapter 14 we will consider the superposition of waves having nearly equal frequencies which leads to the phenomenon of beats.

**Indeed in Sec. 10.2, while discussing stationary waves on a string, we had, at a particular value of x, two sinusoidal waves of the same frequency (but having different initial phases) superposing on each other. However, in general, one could have superposition of displacements which are in different directions; for example, the superposition of two linearly polarized waves to produce a circularly polarized wave (see Chapter 18).

which can be written in the form

$$x(t) = a\cos(\omega t + \theta) \tag{10.5-3}$$

where

$$a\cos\theta = a_1\cos\theta_1 + a_2\cos\theta_2 \tag{10.5-4}$$

and

$$a\sin\theta = a_1\sin\theta_1 + a_2\sin\theta_2 \tag{10.5-5}$$

Thus the resultant disturbance is also simple harmonic in character having the same frequency but different amplitude and different initial phase. If we square and add Eqs. (10.5-4) and (10.5-5), we would obtain

$$a = \left[a_1^2 + a_2^2 + 2a_1a_2\cos(\theta_1 - \theta_2)\right]^{1/2} \tag{10.5-6}$$

Further

$$\tan\theta = \frac{a_1\sin\theta_1 + a_2\sin\theta_2}{a_1\cos\theta_1 + a_2\cos\theta_2} \tag{10.5-7}$$

It should be pointed out that θ is not uniquely determined from Eq. (10.5-7); however, if we assume a to be always positive, then cos θ and sin θ can be determined from Eqs. (10.5-4) and (10.5-5) which will uniquely determine θ.

From Eq. (10.5-6) we find that if

$$\theta_1 \sim \theta_2 = 0,\ 2\pi,\ 4\pi, \ldots \tag{10.5-8}$$

then

$$a = a_1 + a_2 \tag{10.5-9}$$

Thus, if the two displacements are in phase, then the resultant amplitude will be the sum of the two amplitudes; this is known as constructive interference. Similarly, if

$$\theta_1 \sim \theta_2 = \pi,\ 3\pi,\ 5\pi, \ldots \tag{10.5-10}$$

then

$$a = a_1 \sim a_2 \tag{10.5-11}$$

and the resultant amplitude is the difference of the two amplitudes. This is known as destructive interference. If we refer to Fig. 10.2, then we can see that constructive interference occurs at $x = \frac{\lambda}{4}, \frac{3\lambda}{4}, \frac{5\lambda}{4}, \ldots$ (i.e. at the points P', Q', R',...) and destructive interference occurs at $x=0$, $\lambda/2$, λ, $3\lambda/2$...(i.e., at points A, P, Q, R...).* It may be mentioned that when constructive and destructive interferences occur, there is no violation of the principle of conservation of energy; there is merely redistribution of energy.

---

*In the next chapter, we will study the interference pattern produced by the superpositon of spherical waves emanating from two point sources.

In general, if we have n displacements

$$\left.\begin{aligned} x_1 &= a_1 \cos(\omega t+\theta_1) \\ x_2 &= a_2 \cos(\omega t+\theta_2) \\ &\vdots \\ x_n &= a_n \cos(\omega t+\theta_n) \end{aligned}\right\} \tag{10.5-12}$$

then

$$x = x_1+x_2+\ldots+x_n = a\cos(\omega t+\theta) \tag{10.5-13}$$

where

$$a\cos\theta = a_1\cos\theta_1+\ldots+a_n\cos\theta_n \tag{10.5-14}$$

and

$$a\sin\theta = a_1\sin\theta_1+\ldots+a_n\sin\theta_n \tag{10.5-15}$$

10.6 The Graphical Method for Studying Superposition of Sinusoidal Waves

In this section we will discuss the graphical method for adding displacements of the same frequency. This method is particularly useful when we have a large number of superposing waves as it indeed happens when we consider the phenomenon of diffraction.

Let us first try to obtain the resultant of the two displacements given by Eq. (10.5-1) using the graphical method. We draw a circle of radius a_1 and let the point P on the circle be such that OP makes an angle θ_1

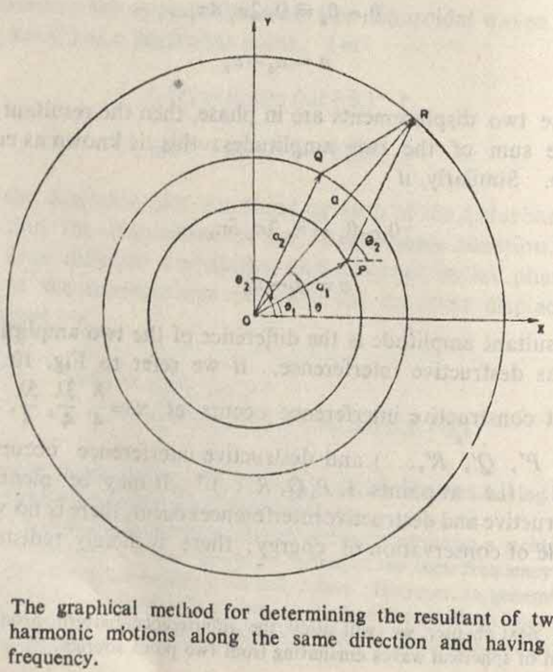


Fig. 10.8 The graphical method for determining the resultant of two simple harmonic motions along the same direction and having the same frequency.

with the x-axis* (see Fig. 10.8). We next draw a circle of radius a_2 and let the point Q on the circle be such that OQ makes an angle θ_2 with the x-axis. We use the law of parallelograms to find the resultant $\overrightarrow{OR}$ of the vectors $\overrightarrow{OP}$ and $\overrightarrow{OQ}$. The length of the vector $\overrightarrow{OR}$ will represent the amplitude of the resultant displacement and if θ is the angle that OR makes with the x-axis, then the initial phase of the resultant will be θ. This can be easily seen by noting that

$$\begin{aligned} OR\cos\theta &= OP\cos\theta_1 + PR\cos\theta_2 \\ &= a_1\cos\theta_1 + a_2\cos\theta_2 \end{aligned} \qquad (10.6\text{-}1)$$

Similarly

$$OR\sin\theta = a_1\sin\theta_1 + a_2\sin\theta_2 \qquad (10.6\text{-}2)$$

consistent with Eqs. (10.5-4) and (10.5-5). Further, as the vectors $\overrightarrow{OP}$ and $\overrightarrow{OQ}$ rotate on the circumference of the circles of radii a_1 and a_2, the vector $\overrightarrow{OR}$ rotates on the circumference of the circle of radius OR with the same frequency.

Thus, if we wish to find the resultant of the two displacements given by Eq. (10.5-1) then we must first draw a vector ($\overrightarrow{OP}$) of length a_1 making an angle θ_1 with the axis; from the tip of this vector we must draw another vector ($\overrightarrow{PR}$) of length a_2 making an angle θ_2 with the axis. The length of the vector $\overrightarrow{OR}$ will represent the resultant amplitude and the angle that it makes with the axis will represent the initial phase of the resultant displacement. It can be easily seen that if we have a third displacement

$$x_3 = a_3\cos(\omega t + \theta_3) \qquad (10.6\text{-}3)$$

then from the point R we must draw a vector $\overrightarrow{RR'}$ of length a_3 which makes an angle θ_3 with the axis; the vector $\overrightarrow{OR'}$ will represent the resultant of x_1, x_2 and x_3.

As an illustration of the above procedure we consider the resultant of N simple harmonic motions all having the same amplitude and with their phases increasing in arithmetic progression. Thus

$$\begin{aligned} x_1 &= a\cos\omega t \\ x_2 &= a\cos[\omega t + \theta_0] \\ \vdots\ & \qquad \vdots \\ x_N &= a\cos[\omega t + (N-1)\theta_0] \end{aligned} \qquad (10.6\text{-}4)$$

---

*Clearly, if we assume the vector $\overrightarrow{OP}$ to rotate (in the anticlockwise direction) with angular velocity ω then the x-coordinate of the vector $\overrightarrow{OP}$ will be $a_1\cos(\omega t + \theta_1)$ where $t=0$ corresponds to the instant when the rotating vector is at the point P.

In Fig. 10.9 the vectors $\overrightarrow{OP_1}$, $\overrightarrow{P_1P_2}$, $\overrightarrow{P_2P_3}$, ... correspond to $x_1, x_2, x_3, \ldots$ respectively. The resultant is denoted by the vector $\overrightarrow{OP_N}$. Let Q_1L and Q_2L be the perpendicular bisectors of OP_1 and P_1P_2. It is easy to prove that

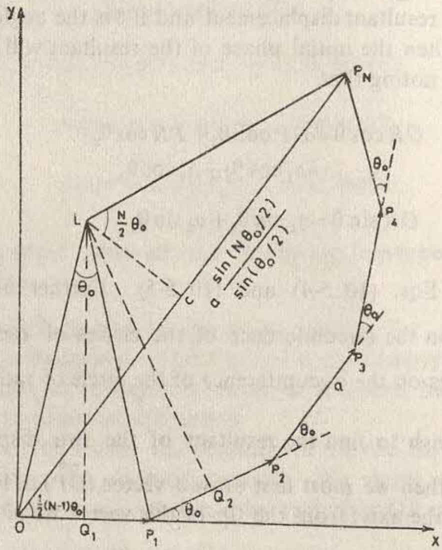


Fig. 10.9 The graphical method for determining the resultant of N simple harmonic motions along the same direction and having the same frequency.

$$\Delta\, LQ_1P_1 \equiv \Delta\, LQ_2P_1$$

Thus $LO=LP_1=LP_2$. Therefore, the points $O, P_1, P_2, P_3, \ldots P_N$ will lie on the circumference of a circle whose centre is L and radius is LO. Further, $\angle LP_1O=\frac{\pi-\theta_0}{2}$ and, therefore, $\angle OLP_1=\theta_0$. Thus

$$LO=\frac{a/2}{\sin\theta_0/2}$$

and

$$OP_N=2OC=2LO\sin\frac{N\theta_0}{2}$$

$$=a\,\frac{\sin\dfrac{N\theta_0}{2}}{\sin\dfrac{\theta_0}{2}} \qquad (10.6\text{-}5)$$

Further, the phase of the resultant displacement would be

$$\angle P_NOX=\tfrac{1}{2}(N-1)\,\theta_0$$

Thus

$$a\cos\omega t+a\cos(\omega t+\theta_0)+\ldots+a\cos[\omega t+(N-1)\theta_0] = A\cos(\omega t+\theta) \tag{10.6-6}$$

where

$$A=\frac{a\sin\dfrac{N\theta_0}{2}}{\sin\dfrac{\theta_0}{2}} \tag{10.6-7}$$

and

$$\theta=\tfrac{1}{2}(N-1)\theta_0 \tag{10.6-8}$$

We will use this result in Chapter 15.

10.7 The Complex Representation

Often it is more convenient to use the complex representation in which the displacement

$$x_1=a_1\cos(\omega t+\theta_1) \tag{10.7-1}$$

is written as

$$x_1=a_1e^{i(\omega t+\theta_1)} \tag{10.7-2}$$

where it is implied that the actual displacement is the real part of x_1. Further, if

$$x_2=a_2e^{i(\omega t+\theta_2)}$$

then

$$\begin{aligned} x_1+x_2 &= (a_1e^{i\theta_1}+a_2e^{i\theta_2})\,e^{i\omega t} \\ &= ae^{i(\omega t+\theta)} \end{aligned} \tag{10.7-3}$$

where

$$ae^{i\theta}=a_1e^{i\theta_1}+a_2e^{i\theta_2} \tag{10.7-4}$$

If we equate the real and imaginary parts of the above equation, we would obtain Eqs. (10.5-4) and (10.5-5).

An interesting illustration of the usefulness of this method is to consider the resultant of the N displacements described by Eq. (10.6-4). Thus we write

$$x_1=ae^{i\omega t},\; x_2=ae^{i(\omega t+\theta_0)},\ldots$$

Hence

$$\begin{aligned} x &= x_1+x_2+\ldots \\ &= ae^{i\omega t}\left[1+e^{i\theta_0}+e^{2i\theta_0}+\ldots+e^{i(N-1)\theta_0}\right] \\ &= ae^{i\omega t}\,\frac{1-e^{Ni\theta_0}}{1-e^{i\theta_0}} = ae^{i\omega t}\,\frac{e^{iN\theta_0/2}}{e^{i\theta_0/2}}\cdot\frac{e^{iN\theta_0/2}-e^{-iN\theta_0/2}}{e^{i\theta_0/2}-e^{-i\theta_0/2}} \\ &= \frac{a\sin\dfrac{N\theta_0}{2}}{\sin\dfrac{\theta_0}{2}}\exp\left[i\left\{\omega t+(N-1)\frac{\theta_0}{2}\right\}\right] \end{aligned} \tag{10.7 5}$$

which is consistent with Eq. (10.6-6). The complex representation is also very useful in considering the spreading of a wave packet (see Sec. 7.5).

It may be noted that whereas

$$\mathrm{Re}\,(x_1)+\mathrm{Re}\,(x_2)=\mathrm{Re}\,(x_1+x_2)$$

but

$$(\mathrm{Re}\,x_1)\,(\mathrm{Re}\,x_2)\neq\mathrm{Re}\,(x_1x_2)$$

where Re(. . .) denotes the "real part of" the quantity inside the brackets. Thus, one must be careful in calculating the intensity of a wave which is proportional to the square of the amplitude. While using the complex representation, one must calculate the amplitude first and then the intensity.

PROBLEMS

1. Standing waves are formed on a stretched string under tension of 1 Newton. The length of the string is 30 cm and it vibrates in 3 loops. If the mass per unit length of the wire is 10 mg/cm, calculate the frequency of the vibrations.
2. In the above problem, if the string is made to vibrate in its fundamental mode, what will be the frequency of vibration?
3. In the experimental arrangement of Wiener, what should be the angle between the film and the mirror if the distance between two consecutive dark bands is 7×10^{-3} cm. Assume $\lambda=6\times10^{-5}$ cm. (Ans. $\sim 1/4°$).
4. Standing waves with five loops are produced on a stretched string under tension. The length of the string is 50 cm and the frequency of vibrations is 250 sec^{-1}. Calculate the time variation of the displacement of the points which are at distances of 2 cm, 5 cm, 15 cm, 18 cm, 20 cm, 35 cm and 45 cm from one end of the string.
5. The displacements associated with two waves (propagating in the same direction) having same amplitude but slightly different frequencies can be written in the form

$$a\cos 2\pi\left(\nu t-\frac{x}{\lambda}\right)\qquad\text{and}\qquad a\cos 2\pi\left[(\nu+\Delta\nu)\,t-\frac{x}{(\lambda-\Delta\lambda)}\right]$$

   (Such displacements are indeed obtained when we have two tuning forks with slightly different frequencies). Discuss the superposition of the displacements and show that at a particular value of x, the intensity will vary with time.
6. In the above problem assume $v=330$ m/sec, $\nu=256\ \mathrm{sec}^{-1}$, $\Delta\nu=2\ \mathrm{sec}^{-1}$ and $a=0.1$ cm. Plot the time variation of the intensity at $x=0,\ \frac{\lambda}{4}$ and $\frac{\lambda}{2}$.
7. Use the complex representation to study the time variation of the resultant displacement at $x=0$ in Problems 5 and 6.
8. Discuss the superposition of two plane waves (of the same frequency and propagating in the same direction) as a function of the phase difference between them. (Such a situation indeed arises when a plane wave gets reflected at the upper and lower surfaces of a glass slab; see Sec. 12.2).
9. In Example 8.1 we had discussed the propagation of a semicircular pulse on a string. Consider two semicircular pulses propagating in opposite directions. At $t=0$, the displacement associated with the pulses propagating in the $+x$ and in the $-x$ directions are given by

$$[R^2-x^2]^{1/2} \quad \text{and} \quad -[R^2-(x-10R)^2]^{1/2}$$

respectively. Plot the resultant disturbance at $t=R/v$, $2.5R/v$, $5R/v$, $7.5R/v$ and $10R/v$; where v denotes the speed of propagation of the wave.

REFERENCES AND SUGGESTED READING

See at the end of Chapter 13

11 Two Beam Interference by Division of Wavefront

Light + Light does not always give more light,
but may in certain circumstances give darkness.

MAX BORN

11.1 Introduction

In the previous chapter, we had considered the superposition of one-dimensional waves propagating on a string and had shown that there is a variation of energy density along the length of the string due to the interference of the two waves (see Fig. 10.5). In general, whenever two waves superpose, one obtains an intensity distribution which is known as the interference pattern. In this chapter, we will consider the interference pattern produced by waves emanating from two point sources. It may be mentioned that with sound waves the interference pattern can be observed without much difficulty because the two interfering waves maintain a constant phase relationship; this is also the case for microwaves. However, for light waves, due to the very process of emission, one cannot observe interference between the waves from two independent sources,* although the interference does take place (see Sec. 11.4). Thus, one tries to derive the interfering waves from a single wave so that the phase relationship is maintained. The methods to achieve this can be classified under two broad categories. Under the first category, in a typical arrangement, a beam is allowed to fall on two closely spaced holes and the two beams emanating from the holes interfere. This method is known as division of wavefront and will be discussed in detail in this chapter. In the other method, known as division of amplitude, a beam is divided at two or more reflecting surfaces and the reflected beams interfere. This will be discussed in the next chapter. We must, however, emphasize that the present and the following chapters are based on one underlying principle, namely the superposition principle.

---

*It is difficult to observe the interference pattern even with two laser beams unless they are phase locked.

It may be mentioned that it is also possible to observe interference using multiple beams; this is known as multiple beam interferometry and will be discussed in Chapter 13. It will be shown there that multiple beam interferometry offers some unique advantages over two beam interferometry.

11.2 Interference Pattern Produced on the Surface of Water

We consider surface waves emanating from two point sources in a water tank. We may have, for example, two sharp needles vibrating up and down at the points S_1 and S_2 (see Fig. 11.1). Although water waves are not really transverse (see Chapter 8) we will, for the sake of simplicity, assume water waves to produce displacements which are transverse to the direction of propagation.

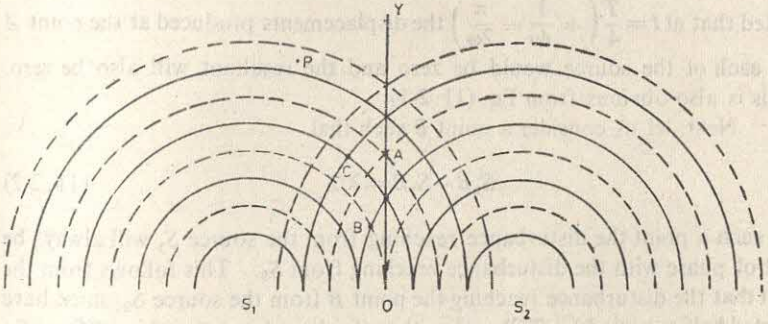


Fig. 11.1 Waves emanating from two point sources S_1 and S_2 vibrating in phase. The solid and the dashed curves represent the positions of the crests and troughs respectively.

If there was only one needle (say at S_1) vibrating with a certain frequency ν then circular ripples would have spread out from the point S_1. The wavelength would have been v/ν and the crests and troughs would have moved outwards. Similarly for the vibrating needle at S_2. However, if both needles are vibrating, then waves emanating from S_1 will interfere with the waves emanating from S_2. We assume that the needle at S_2 vibrates in phase with the needle at S_1; i.e., S_1 and S_2 go up simultaneously, they also reach the lowest position at the same time. Thus, if at a certain instant, the disturbance emanating from the source S_1 produces a crest at a distance ρ from S_1 then the disturbance from S_2 would also produce a crest at a distance ρ from S_2, etc. This is explicitly shown in Fig. 11.1, where the solid curves represent (at a particular instant) the positions of the crests due to disturbances emanating from S_1 and S_2. Similarly, the dashed curves represent (at the same instant) the positions of the troughs. Notice that at all points on the perpendicular bisector OY the disturbances reaching from S_1 and from S_2 will always be in phase. Consequently, at an arbitrary point A (on the perpendicular bisector) we may write the resultant disturbance as

$$\begin{aligned} y &= y_1 + y_2 \\ &= 2a \cos \omega t \end{aligned} \tag{11.2-1}$$

where $y_1 (=a \cos \omega t)$ and $y_2 (=a \cos \omega t)$ represent the displacements at the point A due to S_1 and S_2 respectively. We see that the amplitude at A is twice the amplitude produced by each one of the source. It should be noted that at $t=\frac{T}{4}\left(=\frac{1}{4\nu}=\frac{\pi}{2\omega}\right)$ the displacements produced at the point A by each of the source would be zero and the resultant will also be zero. This is also obvious from Eq. (11.2-1).

Next, let us consider a point B such that

$$S_2B - S_1B = \lambda/2 \tag{11.2-2}$$

At such a point the disturbance reaching from the source S_1 will always be out of phase with the disturbance reaching from S_2. This follows from the fact that the disturbance reaching the point B from the source S_2 must have started half a period $(=T/2)$ earlier than the disturbance reaching B from S_1. Consequently, if the displacement at B due to S_1 is given by

$$y_1 = a \cos \omega t$$

then the displacement at B due to S_2 would be given by

$$y_2 = a \cos (\omega t + \pi) = -a \cos \omega t$$

and the resultant $y = y_1 + y_2$ is zero at all times. Such a point corresponds to destructive interference and is known as a node and corresponds to minimum intensity. It may be pointed out that the amplitudes of the two vibrations reaching the point B will not really be equal as it is at different distances from S_1 and S_2. However, if the distances involved are large (in comparison to the wavelength), the two amplitudes will be very nearly equal and the resultant intensity will be very nearly zero.

In a similar manner we may consider a point C such that

$$S_2C - S_1C = \lambda$$

where the phase of the vibrations (reaching from S_1 and S_2) are exactly the same as at the point A. Consequently we will again have constructive interference.

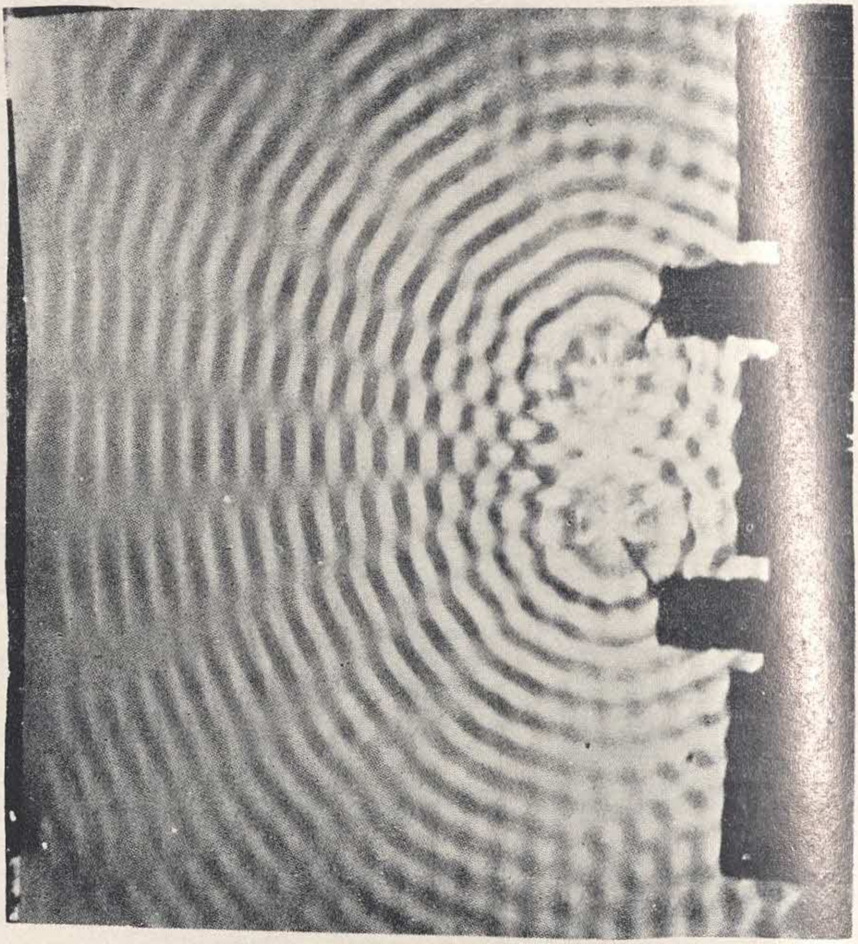


Plate 5 The actual interference pattern produced from two point sources vibrating in phase in a ripple tank. (After PSSC, *Physics*, D.C. Heath & Co., Boston, Mass., 1965; used with permission.)

In general, if a point P is such that

$$S_2P \sim S_1P = n\lambda \qquad \text{(maxima)} \qquad (11.2\text{-}3)$$

$n = 0, 1, 2, \ldots$, then the disturbances reaching the point P from the two sources will be in phase, the interference will be constructive and the intensity will be maximum. On the other hand, if the point P is such that

$$S_2P \sim S_1P = (n+\tfrac{1}{2})\lambda \qquad \text{(minima)} \qquad (11.2\text{-}4)$$

then the disturbances reaching the point P from the two sources will be out of phase, the interference will be destructive and the intensity will be minimum. The actual interference pattern produced from two point sources vibrating in phase in a ripple tank is shown in Plate 5.

Example. 11.1 The intensity at the point which neither satisfies Eq. (11.2-3) nor Eq. (11.2-4) will neither be a maximum nor zero. Consider a point P such that $S_2P - S_1P = \lambda/3$. Find the ratio of the intensity at the point P to that at a maximum.

Solution. If the disturbance reaching the point P from S_1 is given by

$$y_1 = a \cos \omega t$$

then the disturbance from S_2 would be given by

$$y_2 = a \cos \left(\omega t - \frac{2\pi}{3}\right)$$

because a path difference of $\lambda/3$ corresponds to a phase difference of $\frac{2\pi}{3}$. Thus the resultant displacement would be

$$\begin{aligned} y &= y_1 + y_2 \\ &= a\left[\cos \omega t + \cos\left(\omega t - \frac{2\pi}{3}\right)\right] \\ &= 2a \cos\left(\omega t - \frac{\pi}{3}\right) \cos \pi/3 \\ &= a \cos(\omega t - \pi/3) \end{aligned}$$

The intensity is therefore 1/4th of the intensity at the maxima. In a similar manner one can calculate the intensity at any other point.

Example 11.2 The locus of points which correspond to minima are known as nodal lines. Show that the equation of a nodal line is a hyperbola. Also obtain the locus of points which correspond to maxima.

Solution. For the sake of generality we find the locus of the point P which satisfies the following equation

$$S_1P - S_2P = \Delta \qquad (11.2\text{-}5)$$

Thus, if $\Delta = n\lambda$ we have a maximum and if $\Delta = (n+\frac{1}{2})\lambda$ we have a minimum. We choose the midpoint of S_1S_2 as the origin, with the x-axis along S_1S_2 and the y-axis perpendicular to it (see Fig. 11.2). If the distance between S_1S_2 is d then the coordinates of the points S_1 and S_2 are $\left(-\frac{d}{2}, 0\right)$ and $\left(+\frac{d}{2}, 0\right)$ respectively. Let the coordinates of the point P be (x, y). Then

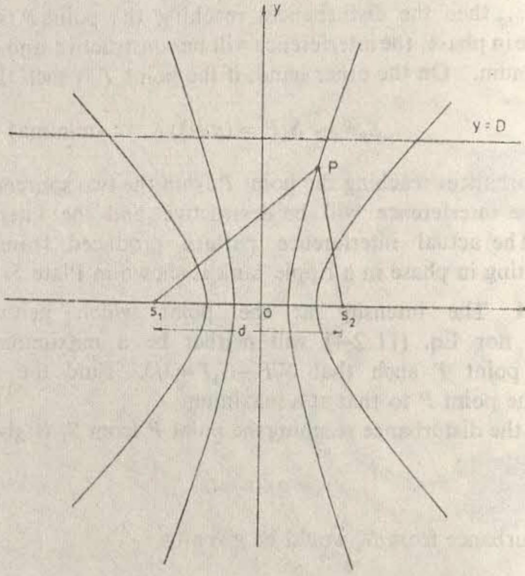


Fig. 11.2 The nodal curves.

$$S_1P = \left[\left(x+\frac{d}{2}\right)^2 + y^2\right]^{1/2}$$

and

$$S_2P = \left[\left(x-\frac{d}{2}\right)^2 + y^2\right]^{1/2}$$

Therefore,

$$S_1P - S_2P = \left[\left(x+\frac{d}{2}\right)^2 + y^2\right]^{1/2} - \left[\left(x-\frac{d}{2}\right)^2 + y^2\right]^{1/2} = \Delta$$

or

$$\left(x+\frac{d}{2}\right)^2 + y^2 = \Delta^2 + \left(x-\frac{d}{2}\right)^2 + y^2 + 2\Delta\left[\left(x-\frac{d}{2}\right)^2 + y^2\right]^{1/2}$$

or

$$2xd - \Delta^2 = 2\Delta\left[\left(x-\frac{d}{2}\right)^2 + y^2\right]^{1/2}$$

On squaring we obtain

$$4x^2d^2 - 4xd\Delta^2 + \Delta^4 = 4\Delta^2\left[x^2 - xd + \frac{d^2}{4} + y^2\right]$$

Thus we obtain

$$\frac{x^2}{\frac{1}{4}\Delta^2}-\frac{y^2}{\frac{1}{4}(d^2-\Delta^2)}=1 \tag{11.2-6}$$

which is the equation of a hyperbola. When $\Delta=(n+\frac{1}{2})\lambda$ the curves correspond to minima and when $\Delta=n\lambda$ the curves correspond to maxima. For large values of x and y the curves asymptotically tend to the straight lines

$$y=\pm\sqrt{\frac{d^2-\Delta^2}{\Delta^2}}\;x \tag{11.2-7}$$

It may be pointed out that there is no point P for which $S_1P\sim S_2P > d$ ($S_1P\sim S_2P$ equals d on the x-axis only). Now, it appears from Eq. (11.2-6) that when $\Delta > d$, the resulting equation is an ellipse which we know is impossible. The fallacy is a result of the fact that because of a few squaring operations, Eq. (11.2-6) also represents the locus of all those points for which $S_1P+S_2P=\Delta$ and obviously in this case Δ can exceed d.

Example 11.3 Consider a line parallel to the x-axis at a distance D from the origin (see Fig. 11.2). Assume $D>>>\lambda$. Find the points on this line where minimum intensity will occur.

Solution. The equation of this line would be

$$y=D \tag{11.2-8}$$

Further at large distances from the origin the equation of the nodal lines would be

$$y=\pm\sqrt{\frac{d^2-\Delta_n^2}{\Delta_n^2}}\;x \tag{11.2-9}$$

where $\Delta_n=(n+\frac{1}{2})\lambda;\ n=0, 1, 2,\ldots$ Clearly the points at which minima will occur (on the line $y=D$) would be given by

$$x_n=\pm\sqrt{\frac{\Delta_n^2}{d^2-\Delta_n^2}}\;D$$

$$=\pm\frac{\Delta_n}{d}\left(1-\frac{\Delta_n^2}{d^2}\right)^{-1/2} D$$

$$\approx\pm(n+\tfrac{1}{2})\frac{\lambda D}{d} \tag{11.2-10}$$

where we have assumed $\Delta_n \ll d$. Thus the points corresponding to minima will be equally spaced with a spacing of $\lambda D/d$.

Example 11.4 Till now we have assumed the needles at S_1 and S_2 (see Fig. 11.1) to vibrate in phase. Assume now that the needles vibrate with a phase difference of π and obtain the nodal lines. Generalize the result for an arbitrary phase difference between the vibrations of the two needles.

Solution. The two needles S_1 and S_2 vibrate out of phase. Thus if, at any instant, the needle at S_1 produces a crest at a distance R from it then the needle at S_2 would produce a trough at a distance R from S_2. Therefore, at all points on the perpendicular bisector OY (see Fig. 11.3) the two vibrations will be always out of phase and we will have minimum. On the other hand, at the point B which satisfies the equation

$$S_2B-S_1B=\lambda/2$$

the two vibrations will be in phase and we will have maximum. Thus,

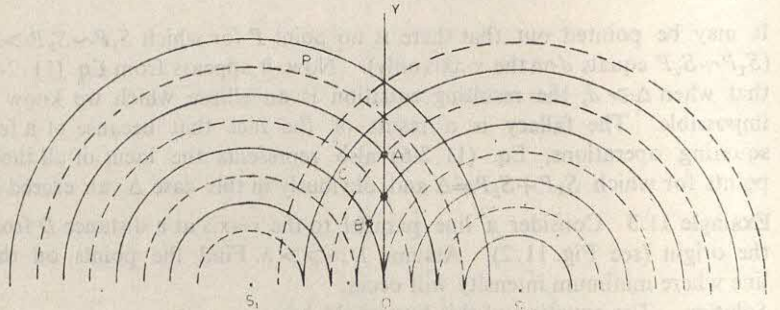


Fig. 11.3 Waves emanating from two point sources S_1 and S_2 vibrating out of phase.

because of the initial phase difference of π, the conditions for maxima and minima are reversed, i.e., when

$$S_2P\sim S_1P=(n+\tfrac{1}{2})\lambda \quad \text{(maxima)}$$

the interference will be constructive and we will have maxima, and when minima

$$S_2P\sim S_1P=n\lambda \quad \text{(minima)}$$

the interference will be destructive and we will have minima. Notice that one again obtains a stationary interference pattern with nodal lines as hyperbolae.

The above analysis can easily be generalized for arbitrary phase difference between the two needles. Assume, for example, that there is a phase difference of $\pi/3$, i.e., if there is a crest at a distance R from S_1 then there is a crest at a distance $R-\lambda/6$ from S_2. Consequently, the condition

$$S_1P-S_2P=n\lambda+\frac{\lambda}{6};\ n=0,\pm 1,\pm 2,\ldots$$

will correspond to maxima.

11.3. Coherence

From the above examples we find that whenever the two needles vibrate with a constant phase difference, a stationary interference pattern is produced. The positions of the maxima and minima will, however, depend on the phase difference in the vibration of two needles. Two sources which vibrate with a fixed phase difference between them are said to be coherent.

Next, assume that the two needles are sometimes vibrating in phase, sometimes vibrating out of phase, sometimes vibrating with a phase difference of $\pi/3$, etc. then the interference pattern will keep on changing. If the phase difference changes with such great rapidity that a stationary interference cannot be observed then the sources are said to be *incoherent*.

Let the displacement produced by the sources at S_1 and S_2 be given by

$$\left.\begin{aligned} y_1 &= a \cos \omega t \\ y_2 &= a \cos (\omega t+\varphi) \end{aligned}\right\} \tag{11.3-1}$$

then the resultant displacement would be

$$y = y_1+y_2 = 2a \cos \varphi/2 \cos (\omega t+\varphi/2) \tag{11.3-2}$$

The intensity (I) which is proportional to the square of the amplitude can be written in the form

$$I = 4I_0 \cos^2 \varphi/2 \tag{11.3-3}$$

where I_0 is the intensity produced by each one of the sources individually. Clearly if $\varphi=\pm\pi, \pm 3\pi, \ldots$, the resultant intensity will be zero and we will have minima. On the other hand, when $\varphi=0, \pm 2\pi, \pm 4\pi, \ldots$ the intensity will be maximum ($=4I_0$). However, if the phase difference between the sources S_1 and S_2 (i.e., φ) is changing with time, the observed intensity will be given by

$$I = 4I_0 \left\langle \cos^2 \frac{\varphi}{2} \right\rangle \tag{11.3-4}$$

where $\langle \ldots \rangle$ denotes the time average of the quantity inside the angular brackets; the time average of a time dependent function is defined by the following relation:

$$\langle f(t) \rangle = \frac{1}{\tau} \int_{-\tau/2}^{+\tau/2} f(t)\, dt \tag{11.3-5}$$

where τ represents the time over which the averaging is carried out. For example, if the interference pattern is viewed by a normal eye this averaging will be over about 1/10th of a second; for a camera with exposure time 0.001 sec, $\tau = 0.001$ sec etc. Clearly, if φ varies in a random manner in times which are small compared to τ then $\cos^2 \varphi/2$ will randomly vary between 0 and 1 and $<\cos^2 \frac{\varphi}{2}>$ would be 1/2. For such a case

$$I=2I_0 \tag{11.3-6}$$

which implies that if the sources are incoherent then the resultant intensity is the sum of the two intensities and there is no variation of intensity ! Thus, if one (or both) of the two vibrating sources are turned on and off in a random manner (such that the phase difference between the vibrations of the two sources varies rapidly) then the interference phenomenon will not be observed. We will discuss this point again in Sec. 11.6 and also in Chapter 14.

11.4 Interference of Light Waves

Till now we have considered interference of waves produced on the surface of water. We will now discuss the interference pattern produced by light waves; however, for light waves it is difficult to observe a stationary interference pattern. For example, if we use two conventional light sources (like two sodium lamps) illuminating two pinholes (see Fig. 11.4), we will not observe any interference pattern on the screen. This can be understood from the following reasoning: In a convential light source, light comes from a large number of independent atoms; each atom emitting light for about 10^{-9} sec, i.e. light emitted by an atom is essentially a pulse lasting for only 10^{-9} sec.* Even if the atoms were emitting under similar conditions, waves from different atoms would differ in their initial phases. Consequently, light coming out from the holes S_1 and S_2 will have a fixed phase relationship for a period of about 10^{-9} sec, hence the interference pattern will keep on changing every billionth of a second. The eye can notice intensity changes which last at least for a tenth of a second and hence we will observe a uniform intensity over the screen. However, if we have a camera whose time of shutter opening can be made less than 10^{-9} sec, then the film will record an interference pattern.** We summarize the above

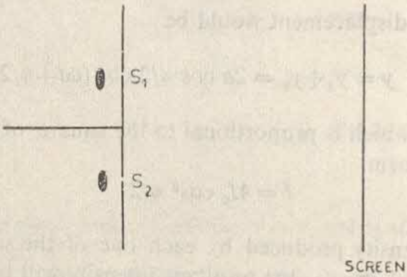


Fig. 11.4 If two sodium lamps illuminate two pinholes S_1 and S_2, no interference pattern will be observed on the screen.

---

*Since the optical frequencies are of the order of 10^{15} sec^{-1}, such a short pulse consists of about a million oscillations; thus it is almost monochromatic (see Chapter 14).

**This interference pattern will be a set of dark and bright bands only if the light waves have the same state of polarization. This can, however, be easily done by putting two polaroids in front of S_1 and S_2. We should mention here that by using

(*Contd.*)

results by noting that light beams from two independent sources do not have any fixed relationship,* as such they do not produce any stationary interference pattern.

Young in 1802 devised an ingenious but simple method to lock the phase relationship between the two sources. The trick lies in the division of a single wavefront into two; these two split wavefronts act as if they emanated from two sources having a fixed phase relationship and, therefore, when these two waves were allowed to interfere, a stationary interference pattern was obtained. In the actual experiment a light source illuminates the pinhole S (see Fig. 11.5). Light diverging from this pinhole fell on a barrier which contained two pinholes S_1 and S_2 which were very close to one another and were located equidistant from S. Spherical waves emanating

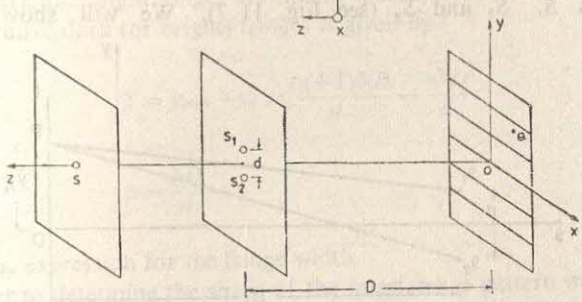


Fig. 11.5 Young's arrangement to produce interference pattern.

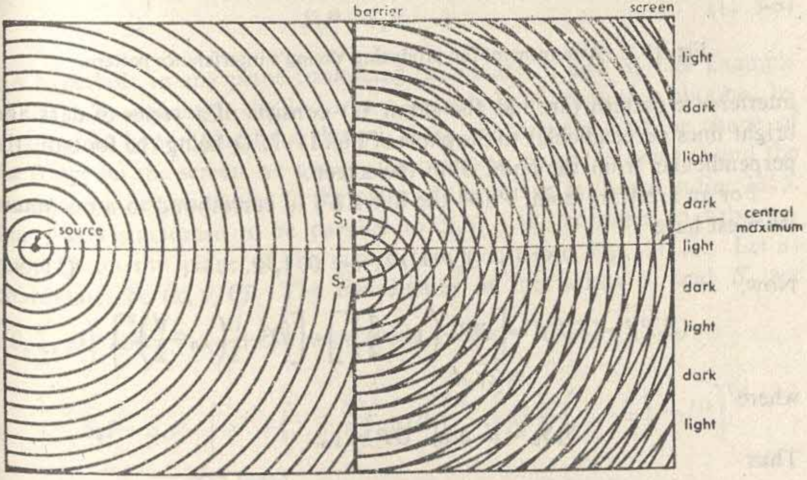


Fig. 11.6 Sections of the spherical wavefronts emanating from S, S_1 and S_2. [Adapted from PSSC Physics (1965).]

---

two independent laser beams it has been possible to record the interference (see Chapter 14).

*Such sources are termed as incoherent sources.

from S_1 and S_2 (see Fig. 11.6) were coherent and on the screen beautiful interference fringes were obtained. In order to show that this was indeed an interference effect, Young showed that the fringes on the screen disappear when S_1 (or S_2) is covered up. Young explained the interference pattern by considering the principle of superposition, and by measuring the distance between the fringes he calculated the wavelength. Fig. 11.6 shows the section of the wavefront on the plane containing S, S_1 and S_2.

11.5 The Interference Pattern

Let S_1 and S_2 represent the two pinholes of the Young's interference experiment. We would determine the positions of maxima and of minima on the line LL' which is parallel to the y-axis and lies in the plane containing the points S, S_1 and S_2 (see Fig. 11.7). We will show that the interference pattern (around the point O) consists of a series of dark and bright lines perpendicular to the plane of Fig. 11.7; O being the foot of the perpendicular from the point S on the screen.

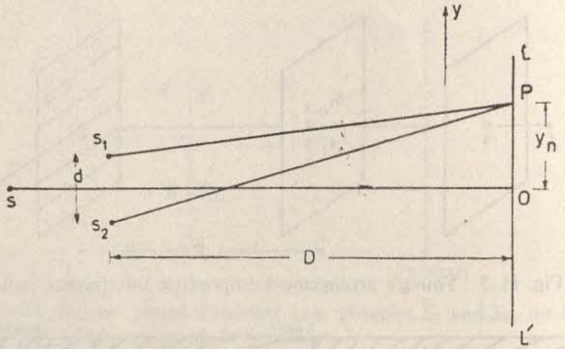


Fig. 11.7 Arrangement for producing Young's interference pattern.

For an arbitrary point P (on the line LL') to correspond to a maximum we must have

$$S_2P - S_1P = n\lambda;\ n = 0, 1, 2, \ldots \tag{11.5-1}$$

Now,

$$(S_2P)^2-(S_1P)^2=\left[D^2+\left(y_n+\frac{d}{2}\right)^2\right]-\left[D^2+\left(y_n-\frac{d}{2}\right)^2\right]$$
$$=2y_nd$$

where

$$S_1S_2=d \text{ and } OP=y_n.$$

Thus

$$S_2P-S_1P=\frac{2y_nd}{S_2P+S_1P} \tag{11.5-2}$$

If $y_n, d \ll D$ then negligible error will be introduced if S_2P+S_1P is replaced by $2D$. For example, for $d=0.02$ cm, $D=50$ cm, $OP=0.5$ cm (which corresponds to typical values for a light interference experiment)

$$S_2P+S_1P=[(50)^2+(0.51)^2]^{1/2}+[(50)^2+(0.49)^2]^{1/2}$$
$$\approx 100.005 \text{ cm.}$$

Thus if we replace S_2P+S_1P by $2D$, the error involved is about 0.005%. In this approximation, Eq. (11.5-2) becomes

$$S_2P-S_1P \approx \frac{y_n d}{D} \tag{11.5-3}$$

Using Eq. (11.5-1) we obtain

$$y_n=\frac{n\lambda D}{d} \tag{11.5-4}$$

Thus the dark and bright fringes are equally spaced and the distance between two consecutive dark (or bright) fringes is given by

$$\beta=y_{n+1}-y_n=\frac{(n+1)\lambda D}{d}-\frac{n\lambda D}{d}$$

or,

$$\beta=\frac{\lambda D}{d} \tag{11.5-5}$$

which is the expression for the fringe width.

In order to determine the shape of the interference pattern we first note that the locus of the point P such that

$$S_2P-S_1P=\Delta \tag{11.5-6}$$

is a hyperbola in any plane containing the points S_1 and S_2 (see Example 11.2). Consequently, the locus is a hyperbola of revolution obtained by rotating the hyperbola about the axis S_1S_2. In order to find the shape of the fringe on the screen we assume the origin to be at the point O and the z-axis to be perpendicular to the plane of the screen as shown in Fig. 11.5. The y-axis is assumed to be parallel to S_2S_1. We consider an arbitrary point Q on the plane of the screen (i.e. $z=0$) (see Fig. 11.5). Let its coordinates be $(x, y, 0)$. The coordinates of the points S_1 and S_2 are $\left(0, \frac{d}{2}, D\right)$ and $\left(0, -\frac{d}{2}, D\right)$ respectively. Thus

$$S_2P-S_1P=\left[x^2+\left(y+\frac{d}{2}\right)^2+D^2\right]^{1/2}-\left[x^2+\left(y-\frac{d}{2}\right)^2+D^2\right]^{1/2}$$
$$=\Delta \text{ (say)}$$

or

$$\left[x^2+\left(y+\frac{d}{2}\right)^2+D^2\right]=\left\{\Delta+\left[x^2+\left(y-\frac{d}{2}\right)^2+D^2\right]^{1/2}\right\}^2$$

or

$$[2yd-\Delta^2]^2=(2\Delta)^2\left[x^2+\left(y-\frac{d}{2}\right)^2+D^2\right]$$

Hence,

$$4x^2(d^2-\Delta^2)-4\Delta^2y^2=\Delta^2(4D^2-\Delta^2)$$

which is the equation of a hyperbola. Thus the shape of the fringes is hyperbolic. On rearranging we get

$$y=\left[\frac{\Delta^2(4D^2-\Delta^2)}{4(d^2-\Delta^2)}\right]^{1/2}\left[1+\frac{4x^2}{4D^2-\Delta^2}\right]^{1/2} \qquad \text{(11.5-}'$$

For values of x such that

$$x^2<<<D^2-\frac{\Delta^2}{4} \qquad (11.5\text{-}8)$$

the loci are straight lines parallel to the x-axis. Thus we obtain straight line fringes on the screen. It should be emphasized that the fringes are straight lines although the sources S_1 and S_2 are point sources. It is easy to see that if we had slits instead of the point sources we would have obtained again straight line fringes with increased intensities.

The fringes so produced are said to be non-localized; they can be photographed by just placing a film on the screen; they can also be seen through an eyepiece.

11.6 The Intensity Distribution

Let $\mathbf{E}_1$ and $\mathbf{E}_2$ be the electric fields produced at the point P by S_1 and S_2 respectively (see Fig. 11.7). The electric fields $\mathbf{E}_1$ and $\mathbf{E}_2$ will, in general, have different directions and different magnitudes. However, if the distances S_1P and S_2P are very large in comparison to the distance S_1S_2, the two fields will almost be in the same direction. Thus, we may write

$$\left.\begin{aligned}\mathbf{E}_1&=\hat{\mathbf{i}}E_{01}\cos\left(\frac{2\pi}{\lambda}S_1P-\omega t\right)\\ \text{and}\qquad \mathbf{E}_2&=\hat{\mathbf{i}}E_{02}\cos\left(\frac{2\pi}{\lambda}S_2P-\omega t\right)\end{aligned}\right\} \qquad (11.6\text{-}1)$$

where $\hat{\mathbf{i}}$ represents the unit vector along the direction of either of the electric fields. The resultant field will be given by

$$\begin{aligned}\mathbf{E}&=\mathbf{E}_1+\mathbf{E}_2\\ &=\hat{\mathbf{i}}\left[E_{01}\cos\left(\frac{2\pi}{\lambda}S_1P-\omega t\right)+E_{02}\cos\left(\frac{2\pi}{\lambda}S_2P-\omega t\right)\right]\end{aligned} \qquad (11.6\text{-}2)$$

The intensity (I) will be proportional to the square of the electric field and will be given by

$$I=KE^2 \qquad (11.6\text{-}3)$$

or
$$I=K\left[E_{01}^2\cos^2\left(\frac{2\pi}{\lambda}S_1P-\omega t\right)+E_{02}^2\cos^2\left(\frac{2\pi}{\lambda}S_2P-\omega t\right)\right.$$
$$\left.+E_{01}E_{02}\left\{\cos\left[\frac{2\pi}{\lambda}(S_2P-S_1P)\right]+\cos\left[2\omega t-\frac{2\pi}{\lambda}(S_2P+S_1P)\right]\right\}\right] \qquad (11.6\text{-}4)$$

where K is a proportionality constant.* For an optical beam the frequency is very large ($\omega \approx 10^{15}$ sec^{-1}) and all the terms depending on ωt will vary with extreme rapidity (10^{15} times in a second); consequently, any detector would record an average value of various quantities. Now

$$\langle\cos^2(\omega t-\theta)\rangle=\frac{1}{2\tau}\int_{-\tau}^{+\tau}\frac{1+\cos 2(\omega t-\theta)}{2}\,dt$$
$$=\tfrac{1}{2}+\frac{1}{16\pi}\frac{T}{\tau}\left\{\Big[\sin 2(\omega t-\theta)\Big]_{-\tau}^{+\tau}\right\}$$

where $T=\frac{2\pi}{\omega}$ ($\approx 2\pi\times 10^{-15}$ sec for an optical beam). For any practical detector** $\frac{T}{\tau} <<< 1$ and since the quantity between the curly brackets will always be between -2 and $+2$ we may write

$$\langle\cos^2(\omega t-\theta)\rangle \approx \tfrac{1}{2} \qquad (11.6\text{-}5)$$

The factor $\cos(2\omega t-\varphi)$ will oscillate between $+1$ and -1 and its average will be zero as can indeed be shown mathematically. Thus the intensity that a detector will record will be given by

$$I=I_1+I_2+2\sqrt{I_1I_2}\cos\delta \qquad (11.6\text{-}6)$$

where

$$\delta=\frac{2\pi}{\lambda}(S_2P-S_1P) \qquad (11.6\text{-}7)$$

represents the phase difference between the displacements reaching the point P from S_1 and S_2. Further

$$I_1=\tfrac{1}{2}KE_{01}^2$$

represents the intensity produced by the source S_1 if no light from S_2 is allowed to fall on the screen; similarly $I_2=\frac{1}{2}KE_{02}^2$ represents the intensity produced by the source S_2 if no light from S_1 is allowed to fall on the screen.

---

*Eq. (11.6-3) will be derived in Chapter 18. In free space the constant K will be shown to be equal to $\epsilon_0 c^2$ where ϵ_0 ($=8.854\times 10^{-12}$ coul^2/N-m^2) represents the permittivity of free space and c, the speed of light in free space.

**For a normal eye, $\tau \approx 0.1$ sec thus $T/\tau \approx 6\times 10^{-14}$; even for a detector having 1 n sec as the resolution time, $T/\tau \approx 6\times 10^{-5}$.

From Eq. (11.6-6) we may deduce the following:

(i) The maximum and minimum values of cos δ are $+1$ and -1 respectively; as such the maximum and minimum values of I are given by

$$\left.\begin{aligned} I_{\max} &= (\sqrt{I_1}+\sqrt{I_2})^2 \\ \text{and} \qquad I_{\min} &= (\sqrt{I_1}-\sqrt{I_2})^2 \end{aligned}\right\} \qquad (11.6\text{-}8$$

The maximum intensity occurs when

$$\delta = 2n\pi; \quad n = 0, 1, 2, \ldots$$

or

$$S_2P \sim S_1P = n\lambda,$$

and the minimum intensity occurs when

$$\delta = (2n+1)\pi; \quad n = 0, 1, 2, \ldots$$

or

$$S_2P \sim S_1P = (n+\tfrac{1}{2})\lambda$$

Notice that when $I_1 = I_2$, the intensity minimum is zero. In general, $I_1 \neq I_2$ and the minimum intensity is not zero.

(ii) If the holes S_1 and S_2 are illuminated by different light sources (see Fig. 11.4), then the phase difference δ will remain constant for about 10^{-9} sec (see discussion in Sec. 11.3) and thus δ would also vary with time* in a random way. If we now carry out the averaging over time scales which are of the order of 10^{-8} sec then

$$< \cos\delta > = 0$$

and we obtain

$$I = I_1 + I_2$$

Thus, for two incoherent sources, the resultant intensity is the sum of the intensities produced by each one of the sources independently and no interference pattern is observed.

(iii) In the arrangement shown in Fig. 11.7, if the distances S_1P and and S_2P are extremely large in comparison to d, then

$$I_1 \approx I_2 = I_0 \,(\text{say})$$

and

$$\begin{aligned} I &= 2I_0 + 2I_0 \cos\delta \\ &= 4I_0 \cos^2 \delta/2 \end{aligned} \qquad (11.6\text{-}9)$$

---

*Notice that this variation occurs in times of the order of 10^{-9} sec which is about a million times longer than the times for variation of the intensity due to the terms depending on ωt. Thus we are justified in first carrying out the averaging which leads to Eq. (11.6-6).

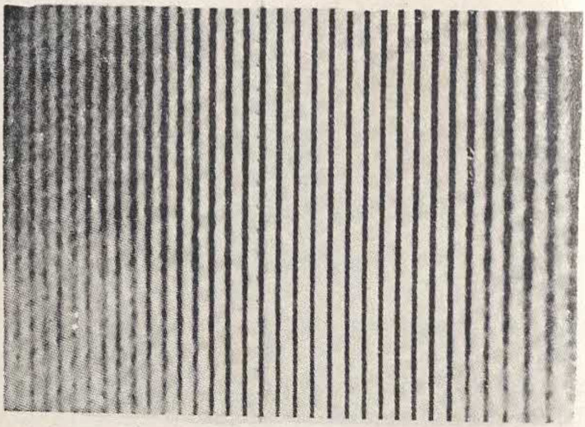


Plate 6 Interference fringes produced by two slits using monochromatic light. (Photograph courtesy Dr. R. S. Sirohi.)

The intensity distribution (which is often termed as the $\cos^2$ pattern) is shown in Fig. 11.8. The actual interference fringes are shown in Plate 6.

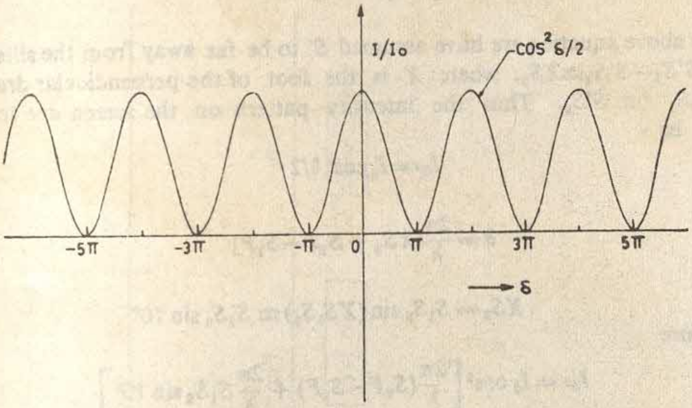


Fig. 11.8 The variation of intensity with δ.

Example 11.5 Consider two parallel slits S_1 and S_2 (perpendicular to the plane of the paper) illuminated by two distant incoherent sources S' and S'' as shown in Fig. 11.9. The angles $S'AO'$ and $S''AO'$ are 10° each. Assuming that both the sources emit almost monochromatic light of the same wavelength λ, determine the intensity pattern on the screen which is at a distance D from the slits (see Fig. 11.9).

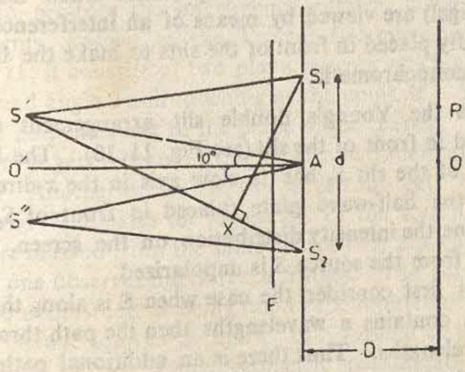


Fig. 11.9 Two distant incoherent sources S' and S'' illuminate the slits S_1 and S_2.

Solution. We will first calculate the intensity pattern produced by each source. Since the two sources are incoherent, the resultant intensity pattern on the screen will be the sum of the intensities produced by each source on the screen.

Consider an arbitrary point P on the screen. Light from the source S' will produce constructive interference at the point P if

$$X S_2+S_2P-S_1P=n\lambda$$

In the above equation we have assumed S' to be far away from the slits so that $S'S_2-S'S_1\approx XS_2$, where X is the foot of the perpendicular drawn from S_1 on $S'S_2$. Thus the intensity pattern on the screen due to S' would be

$$I_{S'}=I_0\cos^2\delta/2$$

where

$$\delta=\frac{2\pi}{\lambda}[XS_2+S_2P-S_1P]$$

But

$$XS_2=S_1S_2\sin(XS_1S_2)\approx S_1S_2\sin 10^\circ$$

Therefore

$$I_{S'}=I_0\cos^2\left[\frac{2\pi}{\lambda}(S_2P-S_1P)+\frac{2\pi}{\lambda}S_1S_2\sin 10^\circ\right]$$

$$=I_0\cos^2\left[\frac{2\pi}{\lambda}\left\{\frac{xd}{D}+d\sin 10^\circ\right\}\right]$$

Similarly

$$I_{S''}=I_0\cos^2\left[\frac{2\pi}{\lambda}\left\{\frac{xd}{D}-d\sin 10^\circ\right\}\right]$$

The resultant intensity would be given by

$$I=I_{S'}+I_{S''}$$

The above example is of practical importance when distant stars (like Betelguse and Rigel) are viewed by means of an interference arrangement. A filter F is usually placed in front of the slits to make the light falling on the slits almost monochromatic.

Example 11.6 In the Young's double slit arrangement two half-wave plates* are placed in front of the slit (see Fig. 11.10). The half-wave plate placed in front of the slit S_1 has its slow axis in the x-direction, whereas the fast axis of the half-wave plate (placed in front of S_2) is along the x-axis. Determine the intensity distribution on the screen. Assume that the light coming from the source S is unpolarized.

Solution. Let us first consider the case when **E** is along the y-axis; if the path through H_1 contains n wavelengths then the path through H_2 would contain $n+\frac{1}{2}$ wavelengths. Thus there is an additional path of $\lambda/2$ introduced in one of the beams and the entire fringe pattern shifts by half of a fringe; i.e., maxima and minima occur when $S_2P\sim S_1P$ is $(n+\frac{1}{2})\lambda$ and $n\lambda$ respectively. A similar fringe pattern is obtained when **E** is in the

---

*Half-wave plates have been introduced in Chapter 18.

y-direction. Consequently, the entire fringe pattern would shift by half of a fringe.

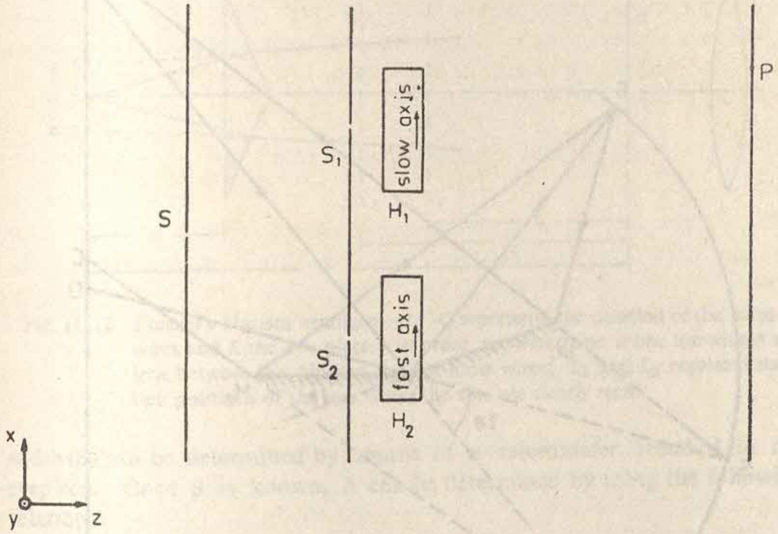


Fig. 11.10 H_1 and H_2 are half-wave plates placed in front of the slits S_1 and S_2.

11.7 Fresnel's Two Mirror Arrangement

After Young's double hole interference experiment, Fresnel devised a series of arrangements to produce the interference pattern. One of the experimental arrangements, known as the Fresnel two-mirror arrangement is shown in Fig. 11.11; it consists of two plane mirrors which are inclined to each other at a small angle θ and touching at the point M. S represents a narrow slit placed perpendicular to the plane of the paper. A portion of the wavefront from S gets reflected from M_1M and illuminates the region AD of the screen. Another portion of the wavefront gets reflected from the mirror MM_2 and illuminates the region BC of the screen. Since these two wavefronts are derived from the same source they are coherent. Thus in the region BC, one observes interference fringes. The formation of the fringes can also be understood as being due to the interference of the wavefronts from the virtual sources S_1 and S_2 of S formed by the mirrors M_1 and M_2 respectively. From simple geometric considerations, it can be shown that the points S, S_1 and S_2 lie on a circle whose centre is at the point M. Further, if the angle between the mirrors is θ, then the angle S_1SS_2 is also θ and the angle S_1MS_2 is 2θ. Thus S_1S_2 is $2R\theta$, where R is the radius of the circle.

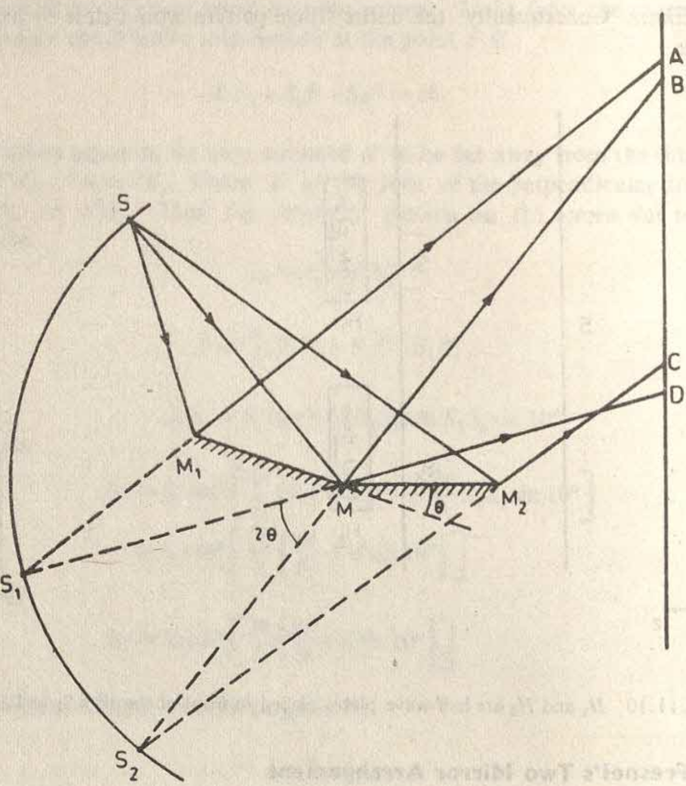


Fig. 11.11 Fresnel's two mirror arrangement.

11.8 Fresnel Biprism

Fresnel devised yet another simple arrangement for the production of interference pattern. He used a biprism, which was actually a simple prism, the base angles of which are extremely small ($\sim\frac{1}{3}°$). The base of the prism is shown in Fig. 11.12 and the prism is assumed to stand perpendicular to the plane of the paper. S represents the slit which is also placed perpendicular to the plane of the paper. Light from the slit S gets refracted by the prism and produces two virtual images S_1 and S_2. These images act as coherent sources and produce interference fringes on the right of the biprism. The fringes can be viewed through an eyepiece. If n represents the refractive index of the material of the biprism and α the base angle then $(n-1)\alpha$ is approximately the angular deviation produced by the prism and, therefore, the distance S_1S_2 is $2a(n-1)\alpha$, where a represents the distance from S to the base of the prism. Thus, for $n=1.5$, $\alpha=\frac{1}{3}° \approx 5.8\times10^{-3}$ radians, $a\approx 2\,\text{cm}$, one gets $d=0.012\,\text{cm}$.

The biprism arrangement can be used for the determination of wavelength of an almost monochromatic light like the one coming from a sodium lamp. Light from the sodium lamp illuminates the slit S and interference fringes can be easily viewed through the eyepiece. The fringe

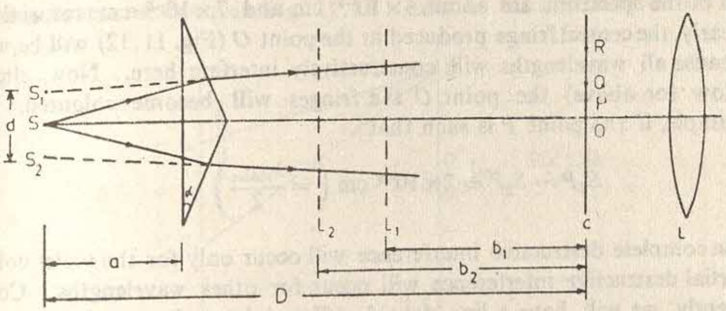


Fig. 11.12 Fresnel's biprism arrangement. C represents the position of the cross-wires and L the eye piece. In order to determine d one introduces a lens between the biprism and the cross wires; L_1 and L_2 represent the two positions of the lens where the slits are clearly seen.

width (β) can be determined by means of a micrometer attached to the eyepiece. Once β is known, λ can be determined by using the following relation:

$$\lambda = \frac{d\beta}{D} \tag{11.8-1}$$

It may be mentioned that in order to determine d, one need not measure the value of α. In fact the distances d and D can easily be determined by placing a convex lens between the biprism and the eyepiece. For a fixed position of the eyepiece there will be two positions of the lens (shown as L_1 and L_2 in Fig. 11.12) where the images of S_1 and S_2 can be seen at the eyepiece.* Let d_1 be the distance between the two images when the lens is at the position L_1 (at a distance b_1 from the eyepiece). Let d_2 and b_2 be the corresponding distances when the lens is at L_2. Then, it can easily be shown that

$$d = \sqrt{d_1 d_2}$$

and

$$D = b_1 + b_2$$

Typically for $d \approx 0.01$ cm, $\lambda \approx 6\times10^{-5}$ cm, $D \approx 50$ cm, $\beta \approx 0.3$ cm.

It should be mentioned that we have considered here a slit instead of a point source. Since, each pair of points S_1 and S_2 produce straight line fringes, the slit will also produce straight line fringes of increased intensity.

---

*This method is similar to the displacement method for the determination of the focal length of a convex lens.

11.9 Interference with White Light

We will now discuss the interference pattern when the slit is illuminated by white light. The wavelengths corresponding to the violet and red end of the spectrum are about 4×10^{-5} cm and 7×10^{-5} cm respectively. Clearly, the central fringe produced at the point O (Fig. 11.12) will be white because all wavelengths will constructively interfere here. Now, slightly below (or above) the point O the fringes will become coloured. For example, if the point P is such that

$$S_2P \sim S_1P = 2\times10^{-5}\ \text{cm}\ \left(=\frac{\lambda_{\text{violet}}}{2}\right)$$

then complete destructive interference will occur only for the violet colour. Partial destructive interference will occur for other wavelengths. Consequently we will have a line devoid of the violet colour and will appear reddish. The point Q which satisfies

$$S_2Q \sim S_1Q = 3.5\times10^{-5}\ \text{cm}\left(=\frac{\lambda_{\text{red}}}{2}\right)$$

will be devoid of the red colour. It will correspond to almost constructive interference for the violet colour. No other wavelength (in the visible region) will neither constructively nor destructively interfere. Thus following the white central fringe we will have coloured fringes; when the path difference is about 2×10^{-5} cm the fringe will be red in colour, then the colour will gradually change to violet. The coloured fringes will soon disappear because at points far away from O there will be so many wavelengths (in the visible region) which will constructively interfere that we will observe uniform white illumination. For example, at a point R, such that $S_2R \sim S_1R = 30\times10^{-5}$ cm wavelengths corresponding to $30\times10^{-5}/n$ $(n=1, 2, \ldots)$ will constructively interfere. In the visible region these wavelengths will be 7.5×10^{-5} cm (red), 6×10^{-5} cm (yellow), 5×10^{-5} cm (greenish yellow) and 4.3×10^{-5} cm (violet). Further, wavelengths corresponding to $30\times10^{-5}/(n+\frac{1}{2})$ will destructively interfere; thus, in the visible region, the wavelengths 6.67×10^{-5} cm (orange), 5.5×10^{-5} cm (yellow), 4.6×10^{-5} cm (indigo) will be absent. The colour of such light, as seen by the unaided eye, will be white. Thus, with white light one gets a white central fringe at the point of zero path difference along with a few coloured fringes on both the sides, the colour soon fading off to white.

It should be noted that in the usual interference pattern with a nearly monochromatic source (like a sodium lamp) a large number of interference fringes are obtained and it is extremely difficult to determine the position of the central fringe. In many interference experiments it is necessary to determine the position of the central fringe and, as has been discussed above, this can easily be done by using white light as a source.

11.10 Displacement of Fringes

We will now discuss the change in the interference pattern produced by introducing a thin transparent plate in the path of one of the two interfering beams as shown in Fig. 11.13. Let t be the thickness of the plate and let n be its refractive index. It is easily seen from the figure that light reaching the point P from S_1 has to traverse a distance t in the plate and a distance S_1P-t in air. Thus the time required for the light to reach from S_1 to the point P is given by

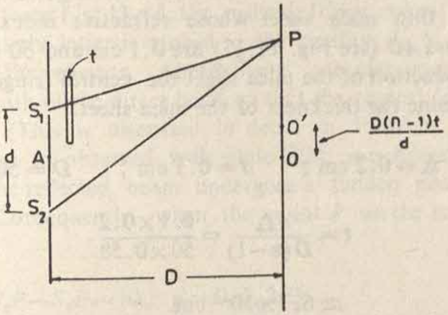


Fig. 11.13 If a thin transparent sheet (of thickness t) is introduced in one of the beams, the fringe pattern gets shifted by a distance $(n-1)\,tD/d$.

$$\frac{S_1P-t}{c}+\frac{t}{v}=\frac{1}{c}\,[S_1P-t+nt]$$

$$=\frac{1}{c}\,[S_1P+(n-1)\,t] \tag{11.10-1}$$

where v $\left(=\frac{c}{n}\right)$ represents the speed of light in the plate. Eq. (11.10-1) shows that by introducing the thin plate the effective optical path increases by $(n-1)\,t$. Thus, when the thin plate is introduced the central fringe (which corresponds to equal optical path from S_1 and S_2) is formed at the point O' where

$$S_1O'+(n-1)\,t=S_2O'$$

Since [see Eq. (11.5-3)]

$$S_2O'-S_1O'\approx\frac{d}{D}\,OO'$$

Therefore

$$(n-1)\,t=\frac{d}{D}\,OO' \tag{11.10-2}$$

Thus the fringe pattern gets shifted by a distance Δ which is given by the following equation

$$\Delta=\frac{D\,(n-1)\,t}{d} \tag{11.10-3}$$

The above principle enables us to determine the thickness of extremely thin transparent sheets (like that of mica) by measuring the displacement of the central fringe. Further, if white light is used as a source, the displacement of the central fringe is easy to measure.

Example 11.7 In a double slit interference arrangement one of the slits is covered by a thin mica sheet whose refractive index is 1.58. The distances S_1S_2 and AO (see Fig. 11.13) are 0.1 cm and 50 cm respectively. Due to the introduction of the mica sheet the central fringe gets shifted by 0.2 cm. Determine the thickness of the mica sheet.

Solution.

$$\Delta = 0.2 \text{ cm} ; \qquad d = 0.1 \text{ cm} ; \qquad D = 50 \text{ cm}$$

Hence

$$t = \frac{d\Delta}{D(n-1)} = \frac{0.1 \times 0.2}{50 \times 0.58}$$

$$\approx 6.7 \times 10^{-4} \text{ cm.}$$

Example 11.8 In an experimental arrangement similar to the one discussed in the above example one finds that by introducing the mica sheet the central fringe occupies the position that was originally occupied by the eleventh bright fringe. If the source of light is a sodium lamp ($\lambda = 5893$ Å) determine the thickness of the mica sheet.

Solution. The point O' (see Fig. 11.13) corresponds to the eleventh bright fringe, thus

$$S_2O' - S_1O' = 11\lambda = (n-1)\,t = 0.58t$$

11.11 The Lloyd's Mirror Arrangement

In this arrangement light from a slit S_1 is allowed to fall on a plane mirror at grazing incidence (see Fig. 11.14). The light directly coming from the slit S_1 interferes with the light reflected from the mirror forming an

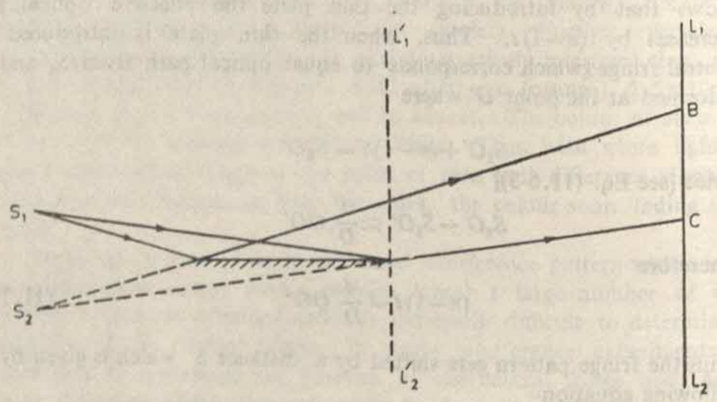


Fig. 11.14 The Lloyd's mirror arrangement.

interference pattern in the region BC of the screen. One may thus consider the slit S_1 and its virtual image S_2 to form two coherent sources which produce the interference pattern. It should be noted that at grazing incidence one really need not have a mirror; even a dielectric surface has very high reflectivity (see Chapter 19).

As can be seen from Fig. 11.14, the central fringe cannot be observed on the screen unless the latter is moved to the position $L_1' L_2'$, where it touches the end of the reflector. Alternatively, one may introduce a thin mica sheet in the path of the direct beam so that the central fringe appears in the region BC. (This is discussed in detail in Problem 2). Indeed, if the central fringe is observed with white light, it is found to be dark. This implies that the reflected beam undergoes a sudden phase change of π on reflection. Consequently, when the point P on the screen is such that

$$S_2P - S_1P = n\lambda, \quad n = 0, 1, 2, 3\ldots$$

we will get minima (i.e. destructive interference). On the other hand, if

$$S_2P - S_1P = (n + \tfrac{1}{2})\,\lambda$$

we will get maxima.

In the next section, using the principle of optical reversibility, we will show that if there is an abrupt phase change of π when light gets reflected by a denser medium, then no such abrupt phase change occurs when reflection takes place at a rarer medium.

11.12 Phase Change on Reflection

We will now investigate the reflection of light at an interface between two media using the principle of optical reversibility. According to this principle, in the absence of any absorption, a light ray that is reflected or refracted will retrace its original path if its direction is reversed.*

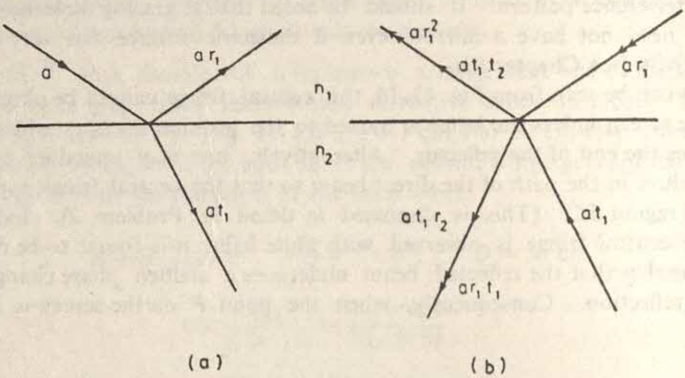
Consider a light ray incident on an interface of two media of refractive indices n_1 and n_2 as shown in Fig. 11.15 (a). Let the amplitude reflection and transmission coefficients be r_1 and t_1 respectively. Thus, if the amplitude of the incident ray is a, then the amplitudes of the reflected and refracted beam would be ar_1 and at_1 respectively.

We now reverse the rays and we consider a ray of amplitude at_1 incident on medium 1 and a ray of amplitude ar_1 incident on medium

---

*This principle is a consequence of time reversal invariance according to which processes can run either way in time. Physicists have yet to find a process which cannot run actually backwards in time if offered the proper circumstances [see, for example, A. Baker, *Modern Physics and Anti Physics*, Chapter 3, Addison Wesley (1970); R. P. Feynman, R. B. Leighton and M. Sands, *The Feynman Lectures on Physics*, Vol. 1. Chapter 52, Addison Wesley (1965)].

2 as shown in Fig. 11.15 (b). The ray of amplitude at_1 will give rise to a reflected ray of amplitude at_1r_2 and a transmitted ray of amplitude



(a) (b)

Fig. 11.15 (a) A ray travelling in a medium of refractive index n_1 incident on a medium of refractive index n_2. (b) Rays of amplitudes ar_1 and at_1 incident on a medium of refractive index n_1.

at_1t_2 where r_2 and t_2 are the amplitude reflection and transmission coefficients when a ray is incident from medium 2 on medium 1. Similarly, the ray of amplitude ar_1 will give rise to a ray of amplitude ar_1^2 and a refracted ray of amplitude ar_1t_1. According to the principle of optical reversibility the two rays of amplitudes ar_1^2 and at_1t_2 must combine to give the incident ray of Fig. 11.15a; thus

$$ar_1^2 + at_1t_2 = a$$

or

$$t_1t_2 = 1 - r_1^2 \tag{11.12-1}$$

Further, the two rays of amplitudes at_1r_2 and ar_1t_1 must cancel each other; i.e.

$$at_1r_2 + ar_1t_1 = 0$$

or

$$r_2 = -r_1 \tag{11.12-2}$$

Since we know from the Lloyd's mirror experiment that an abrupt phase change of π occurs when light gets reflected by a denser medium, we may infer from Eq. (11.12-2) that no such abrupt phase change occurs when light gets reflected by a rarer medium. This is indeed borne out by experiments. Eqs. (11.12-1) and (11.12-2) are known as Stokes' relations.

In Chapter 20, we will calculate the amplitude reflection and transmission coefficients for plane waves incident on a dielectric and also on a conductor. It will be shown that the coefficients satisfy Stokes' relations; the phase change on reflection will also be discussed there.

PROBLEMS

1. In the Young's double-hole experiment (see Fig. 11.5), the distance between the two holes is 0.5 mm, $\lambda \approx 5\times10^{-5}$ cm and $D=50$ cm. What will be the fringe width?
2. Fig. 11.16 represents the layout of Lloyd's mirror experiment. *S* is a point source emitting waves of frequency 6×10^{14} sec^{-1}. *A* and *B* represent the two

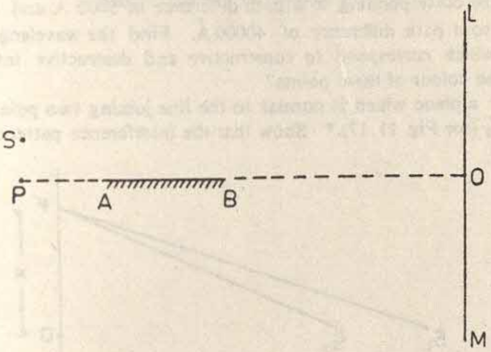


Fig. 11.16 For Problem 2.

ends of a mirror placed horizontally and *LOM* represents the screen. The distances *SP, PA, AB* and *BO* are 1 mm, 5 cm, 5 cm and 190 cm respectively. (a) Determine the position of the region where the fringes will be visible and calculate the number of fringes. (b) Calculate the thickness of a mica sheet ($n=1.5$) which should be introduced in the path of the direct ray so that the lowest fringe becomes the central fringe. The velocity of light is 3×10^{10} cm/sec.

[Ans. (a) 2 cm, 40 fringes, (b) 38 μm].

3. (a) In the Fresnel's biprism arrangement, show that $d=2(n-1)a\alpha$ where *a* represents the distance from the source to the base of the prism (see Fig. 11.12), α is the angle of the biprism and *n* the refractive index of the material of the biprism.

(b) In a typical biprism arrangement $b/a=20$, and for sodium light ($\lambda \simeq 5893$ Å) one obtains a fringe width of 0.1 cm; here *b* is the distance between the biprism and the screen. Assuming $n=1.5$ calculate the angle α.

(Ans. $\simeq 0.71°$)

4. In the Young's double hole experiment a thin mica sheet ($n=1.5$) is introduced in the path of one of the beams. If the central fringe gets shifted by 0.2 cm, calculate the thickness of the mica sheet. Assume $d=0.1$ cm, and $D=50$ cm.
5. In order to determine the distance between the slits in the Fresnel biprism experiment, one puts a convex lens in between the biprism and the eye piece. Show that if $D>4f$ one will obtain two positions of the lens where the image of the slits will be formed at the eye piece; here f is the focal length of the convex lens and *D* is the distance between the slit and the eye piece. If d_1 and d_2 are the distances between the images (of the slits) as measured by the eyepiece, then show that $d=\sqrt{d_1 d_2}$. What would happen if $D<4f$?
6. In the Young's double hole experiment, interference fringes are formed using sodium light which predominantly comprises of two wavelength (5890 Å and 5896 Å). Obtain the regions on the screen where the fringe pattern will disappear. You may assume $d=0.5$ mm and $D=100$ cm.

7. If one carries out the Young's double hole interference experiment using microwaves of wavelength 3 cm, discuss the nature of the fringe pattern if $d=0.1$ cm, 1 cm, and 4 cm. You may assume $D=100$ cm. Can you use Eq. (11.5-5) for the fringe width?
8. In the Fresnel's two mirror arrangement (see Fig. 11.11) show that the points S, S_1 and S_2 lie on a circle and $S_1S_2=2b\theta$ where $b=MS$ and θ is the angle between the mirrors.
9. In the double hole experiment using white light, consider two points on the screen, one corresponding to a path difference of 5000 Å and the other corresponding to a path difference of 40000 Å. Find the wavelengths (in the visible region) which correspond to constructive and destructive interference. What will be the colour of these points?
10. Consider a plane which is normal to the line joining two point coherent sources S_1 and S_2 (see Fig. 11.17).* Show that the interference pattern produced on this plane will consist of concentric circles and that the radius of the n^{th} bright circle will be approximately $\sqrt{2}\sqrt{1-\frac{n\lambda}{d}}\,D$ where we have assumed $D \gg d, x$; the distances d, D and x have been defined in Fig. 11.17. The first bright ring will correspond to the integer which is closest to and smaller than d/λ and as we go outwards, the value of n would decrease.

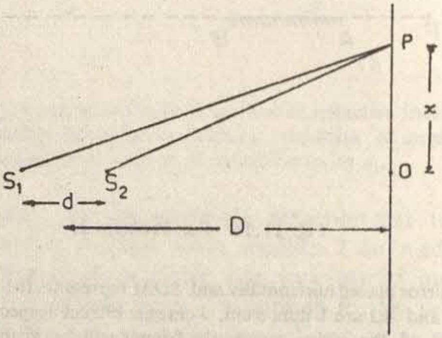


Fig. 11.17 For Problem 10.

11. In the above problem assuming that $d=0.5$ mm, $\lambda=5\times10^{-5}$ cm, $D=100$ cm show that the central fringe is bright and corresponds to $n=1000$. The first, second and third bright rings will correspond to $n=999$, 998 and 997 respectively. Calculate their radii. Also calculate the radii of the first three dark rings.
12. Using the expressions for the amplitude reflection and transmission coefficients (derived in Chapter 20), show that they satisfy Stoke's relations.
13. Assume a plane wave incident normally on a plane containing two holes separated by a distance d. If we place a convex lens behind the slits, show that the fringe width, as observed on the focal plane of the lens will be $f\lambda/d$ where f is the focal length of the lens.
14. In the previous problem, show that if the plane (containing the holes) lies in the front focal plane of the lens, then the interference pattern will consist of exactly parallel straight lines. However, if the plane does not lie on the front focal plane, the fringe pattern will be hyperbolae as assumed in Sec. 11.5.

---

*The arrangement is known as Pohl's interferometer and S_1 and S_2 could be the virtual images of a point source placed in front of a thin mica sheet.

15. In the Young's double hole experiment calculate $I/I_{\max}$ where I represents the intensity at a point where the path difference is $\lambda/5$.

REFERENCES AND SUGGESTED READING

See at the end of Chapter 13

12 Interference by Division of Amplitude

The issue is now quite clear. It is between light and darkness, and everyone must choose his side.

CHESTERTON

12.1 Introduction

In the previous chapter we discussed the interference pattern produced by division of a wavefront; for example, light coming out of a pinhole was allowed to fall on two holes, and spherical waves emanating from these two holes produced the interference pattern. In this chapter we will consider the formation of interference pattern by division of amplitude; for example, if a plane wave falls on a thin film then the wave reflected from the upper surface interferes with the wave reflected from the lower surface. Such studies have many practical applications and also explain phenomena like the formation of beautiful colours produced by a soap film illuminated by white light.

12.2 Interference by a Plane Parallel Film When Illuminated by a Plane Wave

If a plane wave is incident normally on a thin* film of uniform thickness t (see Fig. 12.1) then the waves reflected from the upper surface interfere with the waves reflected from the lower surface; in this section we will study this interference pattern. In order to observe the interference pattern without obstructing the incident beam, we use a partially reflecting plate G as shown in Fig. 12.1. Such an arrangement also enables us to eliminate the direct beam from reaching the photographic plate P (or the

---

*Why the film should be thin is explained in Sec. 12.7.

eye). The plane wave may be produced by placing an illuminated pinhole at the focal point of a corrected lens; alternatively, it may just be a beam coming out of a laser.

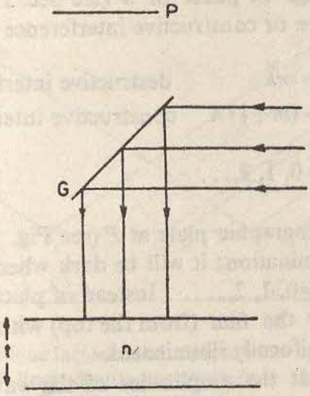


Fig. 12.1 The normal incidence of a parallel beam of light on a thin film of refractive index *n* and thickness *t*. *G* denotes a partially reflecting plate and *P* represents a photographic plate.

Let the solid and the dashed lines in Fig. 12.2 represent the positions of the crests[*] (at any particular instant of time) corresponding to the waves reflected from the upper and lower surfaces of the film respectively.[**] Clearly, the wave reflected from the lower surface of the

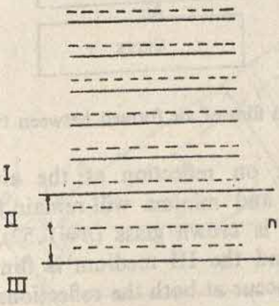


Fig. 12.2 The solid and the dashed lines represent the crests of the waves reflected from the upper surface and from the lower surface of the thin film. Notice that the distance between the consecutive crests inside the film is less than the corresponding distance in medium I.

---

[*] Notice that the distance between consecutive crests in the film is less than the corresponding distance in air. This is because of the fact that the effective wavelength in a medium of refractive index n is λ/n.

[**] In general, the wave reflected from the lower surface of the film will suffer multiple reflections. The effect of such multiple reflection is neglected (see next chapter).

film traverses an additional optical path of $2nt$, where n represents the refractive index of the material of the film. Further, if the film is placed in air then, the wave reflected from the upper surface of the film will undergo a sudden change in phase of π (see Sec. 11.12) and as such the conditions for destructive or constructive interference will be given by

$$2nt = m\lambda \qquad \text{destructive interference} \qquad (12.2\text{-}1a)$$

$$= (m+\tfrac{1}{2})\,\lambda \qquad \text{constructive interference} \qquad (12.2\text{-}1b)$$

where

$$m = 0, 1, 2, \ldots$$

Thus, if we place a photographic plate at P (see Fig. 12.1), then the plate will receive uniform illumination; it will be dark when $2nt=m\lambda$ and bright when $2nt=(m+\frac{1}{2})\,\lambda$; $m=0, 1, 2, \ldots$. Instead of placing the photographic plate, if we try to view the film (from the top) with naked eye, then the film will appear to be uniformly illuminated.

It may be noted that the amplitudes of the waves reflected from the upper and lower surfaces will, in general, be slightly different; and as such the interference will not be completely destructive. However with appropriate choice of the refractive indices of media II and III, the two amplitudes can be made very nearly equal (see Example 12.1).

For an air film between two glass plates (see Fig. 12.3) no phase change will occur on reflection at the glass-air interface, but a phase change of π will occur on reflection at the air-glass interface and the conditions for maxima and minima will remain the same. On the other hand, if the I medium is crown glass ($n=1.52$), II medium is an oil of refractive index 1.60 and the III medium is flint glass ($n=1.66$) then a phase change of π will occur at both the reflections and the conditions for maxima and minima would be

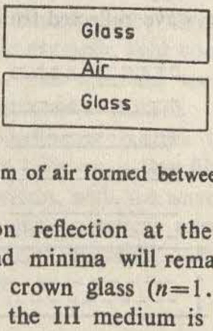


Fig. 12.3 Thin film of air formed between two glass plates.

$$2nt = (m+\tfrac{1}{2})\,\lambda \qquad \text{minima} \qquad (12.2\text{-}2a)$$

$$= m\lambda \qquad \text{maxima} \qquad (12.2\text{-}2b)$$

In general, whenever the refractive index of the II medium lies in between the refractive indices of the I and the III media, then the conditions of maxima and minima would be given by Eq. (12.2-2).

We next consider the oblique incidence of a plane wave on the thin film (see Fig. 12.4). Once again, the wave reflected from the upper surface of the film interferes with the wave reflected from the lower surface of the film. The latter traverses an additional optical path Δ, which is given by (see Fig. 12.5):

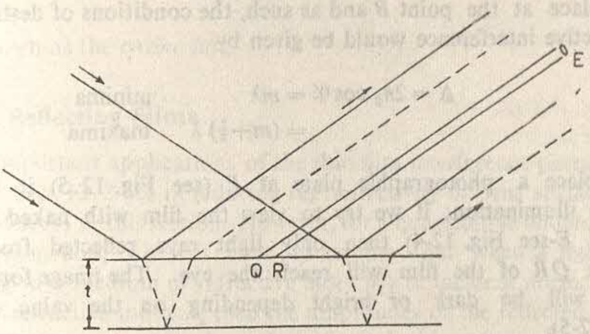


Fig. 12.4 The oblique incidence of a plane wave on a thin film. The solid and dashed lines denote the boundary of the wave reflected from the upper surface and from the lower surface of the film. The eye *E* receives the light reflected from the region *QR*.

$$\Delta = n_2(BD+DF)-n_1BC \qquad (12.2\text{-}3)$$

where *C* is the foot of the perpendicular from the point *F* on *BG*. We will show in the next section that

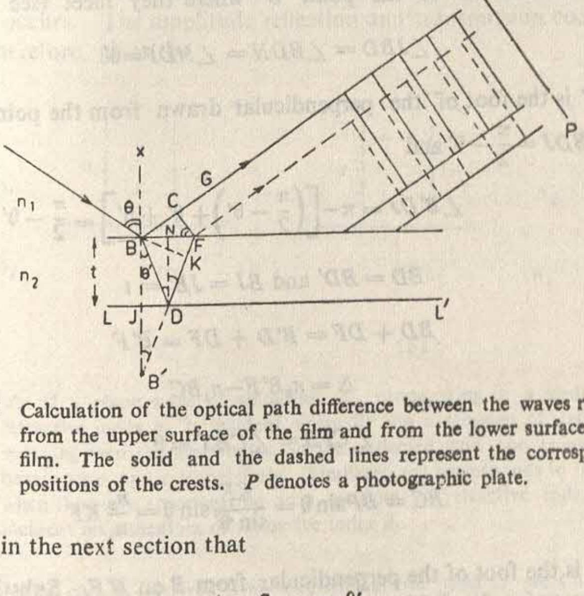


Fig. 12.5 Calculation of the optical path difference between the waves reflected from the upper surface of the film and from the lower surface of the film. The solid and the dashed lines represent the corresponding positions of the crests. *P* denotes a photographic plate.

$$\Delta = 2n_2t\cos\theta' \qquad (12.2\text{-}4)$$

where θ' is the angle of refraction.

For a film placed in air, a phase change of π will occur when reflection takes place at the point B and as such, the conditions of destructive and constructive interference would be given by

$$\Delta = 2n_2 \cos\theta' = m\lambda \qquad \text{minima} \qquad (12.2\text{-}5a)$$

$$= (m+\tfrac{1}{2})\,\lambda \qquad \text{maxima} \qquad (12.2\text{-}5b)$$

If we place a photographic plate at P (see Fig. 12.5) it will receive uniform illumination; if we try to view the film with naked eye (at the position E-see Fig. 12-4) then only light rays reflected from a small position QR of the film will reach the eye. The image formed at the retina will be dark or bright depending on the value of Δ (see Eq. 12.2-5).

12.3 The Cosine Law

In this section we will show that the wave reflected from the lower surface of the film traverses an additional optical path which is given by the following expression:

$$\Delta\,[= n_2(BD+DF)-n_1 BC] = 2n_2 t \cos\theta' \qquad (12.3\text{-}1)$$

Let θ and θ' denote the angles of incidence and refraction respectively. We drop a perpendicular BJ from the point B on the lower surface LL' and extend BJ and FD to the point B' where they meet (see Fig. 12.5). Clearly,

$$\angle JBD = \angle BDN = \angle NDF = \theta'$$

where N is the foot of the perpendicular drawn from the point D on BF.

Now $\angle BDJ = \frac{\pi}{2} - \theta'$ and

$$\angle B'DJ = \pi - \left[\left(\frac{\pi}{2} - \theta'\right) + \theta' + \theta'\right] = \frac{\pi}{2} - \theta'$$

Thus

$$BD = BD' \text{ and } BJ = JB' = t$$

or

$$BD + DF = B'D + DF = B'F$$

Hence

$$\Delta = n_2 B'F - n_1 BC \qquad (12.3\text{-}2)$$

Now

$$\angle CFB = \angle CBX = \theta$$

Thus

$$BC = BF \sin\theta = \frac{KF}{\sin\theta'}\sin\theta = \frac{n_2}{n_1}KF \qquad (12.3\text{-}3)$$

where K is the foot of the perpendicular from B on $B'F$. Substituting the above expression for BC in Eq. (12.3-2) we get

$$\Delta = n_2B'F - n_2KF = n_2B'K$$

or

$$\Delta = 2n_2t\cos\theta' \qquad (12.3\text{-}4)$$

which is known as the *cosine law.*

12.4 Non Reflecting Films

One of the important applications of the thin film interference phenomenon discussed in Sec. 12.2 lies in reducing the reflectivity of lens surfaces; this we plan to discuss in this section. However, for a quantitative understanding of the phenomenon, we will have to assume that when a light beam (propagating in a medium of refractive index n_1) is incident normally on a dielectric of refractive index n_2 then the amplitudes of the reflected and the transmitted beams are related to that of the incident beam through the following relations* (see Fig. 12.6a):

$$a_r = \frac{n_1 - n_2}{n_1 + n_2}\, a_i \qquad (12.4\text{-}1a)$$

$$a_t = \frac{2n_1}{n_1 + n_2}\, a_i \qquad (12.4\text{-}1b)$$

where a_i, a_r and a_t are the amplitudes of the incident beam, reflected beam and the transmitted beam respectively. Notice that when $n_2 > n_1$, a_r is negative showing that when a reflection occurs at a denser medium a phase change of π occurs. The amplitude reflection and transmission coefficients r and t are therefore, given by

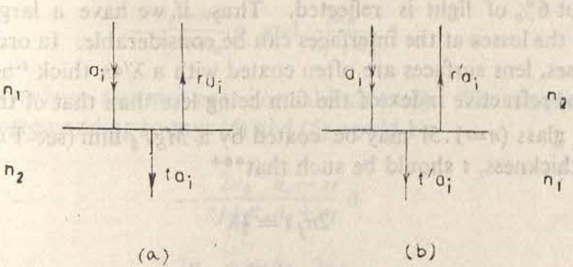


Fig. 12.6 (a) If a plane wave of amplitude a_i, propagating in a medium of refractive index n_1, is incident normally on a medium of refractive index n_2 then the amplitudes of the reflected and the transmitted beams are a_r and a_t respectively. Similarly, (b) corresponds to the case when the beam (propagating in a medium of refractive index n_2) is incident on a medium of refractive index n_1.

---

*These relations can be derived from electromagnetic theory; see Chapter 20 [Eqs. (20.2-69) and (20.2-72)].

$$r=\frac{n_1-n_2}{n_1+n_2} \tag{12.4-2a}$$

$$t=\frac{2n_1}{n_1+n_2} \tag{12.4-2b}$$

It is interesting to point out that if r' and t' are the reflection and transmission coefficients where light (propagating in a medium of refractive index n_2) is incident on a medium of refractive index n_1 [see Fig. 12.6 (b)] then

$$r'=\frac{n_2-n_1}{n_2+n_1}=-r \tag{12.4-3}$$

$$t'=\frac{2n_2}{n_1+n_2} \tag{12.4-4}$$

and

$$1-tt'=1-\frac{4n_1n_2}{(n_1+n_2)^2}=\left(\frac{n_1-n_2}{n_1+n_2}\right)^2=r^2 \tag{12.4-5}$$

Eqs. (12.4-4) and (12.4-5) represent the Stokes' relations (see Sec. 11.12).

We will now discuss the application of the thin film interference phenomenon in reducing the reflectivity of lens surfaces. We all know that in many optical instruments (like a telescope) there are many interfaces and the loss of intensity due to reflections can be severe. For example, for near normal incidence,* the reflectivity of crown glass surface (in air) is**

$$\left(\frac{1.5-1}{1.5+1}\right)^2 \simeq 0.04\,,$$

i.e. 4% of the incident light is reflected. For a dense flint glass $n \simeq 1.67$ and about 6% of light is reflected. Thus, if we have a large number of surfaces, the losses at the interfaces can be considerable. In order to reduce these losses, lens surfaces are often coated with a $\lambda/4n$ thick "non-reflecting film"; the refractive index of the film being less than that of the lens. For example, glass ($n=1.5$) may be coated by a MgF_2 film (see Fig. 12.7) and the film thickness, t should be such that***

$$2n_f\,t=\tfrac{1}{2}\lambda$$

or

$$t=\frac{\lambda}{4n_f} \tag{12.4-6}$$

---

*In all what follows in this section, we will assume near normal incidence.

**The reflectivity is $\left(\frac{n-1}{n+1}\right)^2$

***Since the refractive index of the non-reflecting film is greater than that of the air and less than that of the glass, abrupt phase change of π occurs at both the reflections. Consequently, when $2nt\cos\theta'=m\lambda$ there would be constructive interference and when $2nt\cos\theta'=(m+\frac{1}{2})\lambda$ there would be destructive interference.

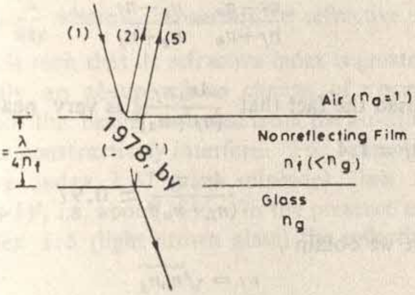


Fig. 12.7 If a film (having a thickness of $\lambda/4n_f$ and having refractive index less than that of the glass) is coated on glass, then waves reflected from the upper surface of the film destructively interfere with the waves reflected from the lower surface of the film. Such a film is known as a non-reflecting film.

where we have assumed near normal incidence (i.e. $\cos\theta' \cong 1$; see Eq. 12.3-4) and n_f represents the refractive index of the film; for MgF_2, $n_f=1.38$. Thus, if we assume λ to be 5.0×10^{-5} cm (which roughly corresponds to the centre of the visible spectrum) we will have

$$t=\frac{5.0\times10^{-5}\text{ cm}}{4\times1.38}\approx 0.9\times10^{-5}\text{ cm}$$

We would like to emphasize the following points:

(i) Let n_a, n_f and n_g be the refractive indices of air, non-reflecting film and glass respectively. If a is the amplitude of the incident wave then the amplitudes of the reflected and refracted waves (the corresponding rays shown as (2) and (3) in Fig. 12.7) would be

$$-\frac{n_f-n_a}{n_f+n_a}a \text{ and } \frac{2n_a}{n_f+n_a}a$$

respectively (we have assumed near normal incidence). The amplitudes of the waves corresponding to rays (4) and (5) would be

$$-\frac{2n_a}{n_f+n_a}\frac{n_g-n_f}{n_g+n_f}a$$

and

$$-\frac{2n_a}{n_f+n_a}\frac{n_g-n_f}{n_g+n_f}\frac{2n_f}{n_f+n_a}a$$

respectively. Now, for complete destructive interference, the waves corresponding to rays (2) and (5) should have the same amplitude, i.e.

$$-\frac{n_f-n_a}{n_f+n_a}a=-\frac{2n_a}{n_f+n_a}\frac{n_g-n_f}{n_g+n_f}\frac{2n_f}{n_f+n_a}a \tag{12.4-7}$$

or

$$\frac{n_f-n_a}{n_f+n_a}=\frac{n_g-n_f}{n_g+n_f} \tag{12.4-8}$$

where we have used the fact that $\frac{4n_a n_f}{(n_f+n_a)^2}$ is very nearly equal to unity; for $n_a=1$ and $n_f=1.4$,

$$\frac{4n_a n_f}{(n_f+n_a)^2}\approx 0.97$$

On simplification we obtain

$$n_f=\sqrt{n_a n_g}$$

If the first medium is air then $n_a=1$ and with $n_g=1.66$ (dense flint glass) n_f should be 1.29 and when $n_g=1.5$ (light crown glass) n_f should be 1.22. We note that the refractive indices of magnesium flouride and cryolite are 1.38 and 1.36 respectively. Now for a $\frac{\lambda}{4n}$ thick film, the reflectivity will be

$$\left[\frac{n_f-n_a}{n_f+n_a}-\frac{n_g-n_f}{n_g+n_f}\right]^2$$

Thus, for $n_a=1, n_f=1.38$ and $n_g=1.5$ the reflectivity will be about 1.3%. In the absence of the film the reflectivity would have been about 4%. The reduction of reflectivity is much more pronounced for the dense flint glass. This technique of reducing the reflectivity is known as *blooming*.

(ii) The film is non-reflecting only for a particular value of λ; in Eq. (12.4-6) λ was assumed to be 5000 Å. For a polychromatic light, the film's non-reflecting property will be falling off when λ is greater or less than the above value. However, the effect is not serious. For example, for the MgF_2 film on crown glass at 5000 Å, the reflectivity rises by about 0.5% as one goes either to the red or the violet end of the visible spectrum. This tells us why we should use a $\frac{\lambda}{4n}$ thick film and not $\frac{3\lambda}{4n}$ or $\frac{5\lambda}{4n}$ thick film, although the latter will also give destructive interference for the chosen wavelength. Thus for $t=2.7\times10^{-5}$ cm (which is three times the value given by Eq. 12.4-6), reflected beams corresponding to $\lambda=7500$ Å will constructively interfere.

(iii) As in the case of Young's double slit experiment there is no loss of energy; there is merely a redistribution of energy. The energy appears mostly in the transmitted beam.

12.5 High Reflectivity by Thin Film Deposition

Another important application of the thin film interference phenomenon is the converse of the procedure just discussed, viz. the glass surface is coated by a thin film of a suitable material to increase the reflectivity. The film thickness is again $\frac{\lambda}{4n_f}$ where n_f represents the refractive index of the film; however, the film is such that its refractive index is greater than that of the glass, consequently an abrupt phase change of π occurs only at the air-film interface and the beams reflected from the air-film interface and the film glass interface constructively interfere. For example, if we consider a film of refractive index 2.37 (zinc sulphide) then the reflectivity is $(2.37-1)^2/(2.37+1)^2$, i.e. about 16%. In the presence of the glass surface of refractive index 1.5 (light crown glass) the reflectivity would become (see Eq. 12.4-10):

$$\left[-\frac{2.37-1}{2.37+1}-\frac{4\times1\times2.37}{(3.37)^2}\times\frac{2.37-1.5}{2.37+1.5}\right]^2$$

which gives over 30%. It should be noted that if the difference between the refractive indices of the film and the glass is increased the reflectivity would also increase.

12.6 Interference by a Plane Parallel Film When Illuminated by a Point Source

In Sec. 12.2 we had considered the incidence of a parallel beam of light on a thin film and had discussed the interference produced by the waves reflected from the upper and lower surfaces of the film. We will now consider the illumination of the film by a point source of light and, once again, in order to observe the film without obstructing the incident beam, we will use a partially reflecting plate G as shown in Fig. 12.8. However,

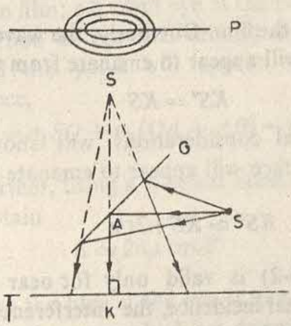


Fig. 12.8 Light emanating from a point source S is allowed to fall on a thin film of thickness t. G is a partially reflecting plate and P represents the photographic plate. On the photographic plate circular fringes are obtained.

in order to study the interference pattern we may *assume* the point source S to be right above the film (see Fig. 12.8) such that the distance SK (in Fig. 12.9) is equal to $SA+AK'$ (in Fig. 12.8); $K'A$ (in Fig. 12.8) and KS (in Fig. 12.9) being normal to the film. Obviously, the waves reflected from the upper surface of the film will appear to emanate from the point S' where

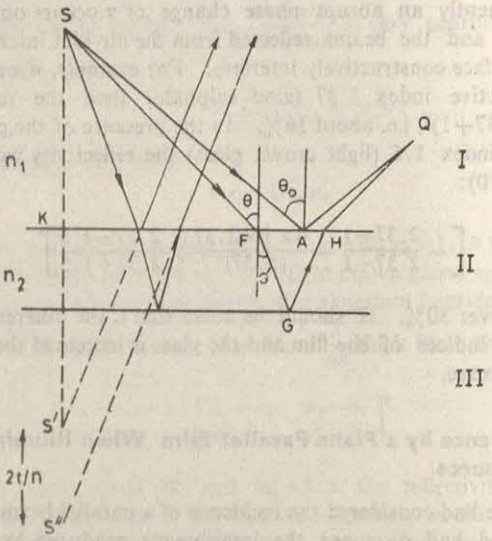


Fig. 12.9 If light emanating from a point source S is incident on a thin film then the interference pattern produced in region I is approximately the same as would have been produced by two coherent point sources S' and $S''\left(\text{separated by a distance } \frac{2t}{n}\right)$ where t represents the thickness of the film and n represents the refractive index of the film.

$$KS' = KS \tag{12.6-1}$$

Further, simple geometrical considerations will show that the waves reflected from the lower surface will appear to emanate from the point S'', where

$$KS'' \simeq KS + 2t/n \tag{12.6-2}$$

(see Fig. 12.9). Eq. (12.6-2) is valid only for near normal incidence.* Thus, at least for near normal incidence, the interference pattern produced in region I (see Fig. 12.9) will be very nearly** the same as produced by two point coherent sources S' and S'' (which is the double hole experiment of Young discussed in the previous chapter). Thus, if we put a photographic plate P (see Fig. 12.8), we will in general, obtain interference fringes. The intensity of an arbitrary point Q will be determined by the following relations:

---

*This is a consequence of the fact that the image of a point source produced by a plane refracting surface is not perfect.

**The fact that this is not identical to the Young's pattern is because of the fact that S'' is not a perfect image of the point S. For large angles of incidence, the waves reflected from the lower surface will appear to emanate from a point which will be displaced from S''.

$$\Delta = (m+\tfrac{1}{2})\lambda \qquad \text{Maxima} \qquad (12.6\text{-}3a)$$

$$= m\lambda \qquad \text{Minima} \qquad (12.6\text{-}3b)$$

where

$$\Delta = [n_1 SF + n_2(FG + GH) + n_1 HQ] - [n_1(SA + AQ)] \qquad (12.6\text{-}4)$$

represents the optical path difference and we have assumed that in one of the reflections, an abrupt phase change of π occurs; n_1 and n_2 are the refractive indices of media I and II respectively. The above conditions are rigorously correct; i.e. valid even for large angles of incidence. Further, it can be shown that for near normal incidence

$$\Delta \simeq 2n_2 t\cos\theta' \qquad (12.6\text{-}5)$$

A more rigorous calculation shows

$$\Delta \simeq 2n_2 t\cos\theta'\left[1 - \frac{n_1^2\sin\theta\cos\theta}{n_2^2 - n_1^2\sin^2\theta}\left(\frac{\theta_0-\theta}{2}\right)\right] \qquad (12.6\text{-}6)$$

where the angles θ, θ_0 and θ' are defined in Fig. 12.9.

Now, if we put a photographic plate (parallel to the surface of the film—see Fig. 12.8) we will obtain dark and bright concentric rings (see Problem 10 in Chapter 11).[*] On the other hand, if we view the film with naked eye then, for a given position of the eye, we will be able to see only a very small portion of the film; e.g. with eye at the position E and the point source at S only a portion of the film around the point B will be visible [see Fig. 12·10(a)], and this point will appear to be dark or bright as the optical path difference,

$$\Delta = n_1 SQ + n_2(QA + AB) - n_1 SB$$

is $m\lambda$ or $(m+\frac{1}{2})\lambda$. Further, using a method similar to the one described in Sec. 12.3, we can obtain

$$\Delta \simeq 2n_2 t\cos\theta' \qquad (12.6\text{-}7)$$

Instead of looking at the film, if the eye is focussed at infinity then the interference is between the rays which are derived from a single incident ray by reflection from the upper and the lower surfaces of the film [see Fig. 12.10(b)]. For example, the rays PM and QR, which focus at the point O

---

[*] If the point source is taken far away then, it can easily be seen that the rings will spread out and in the limit of the point source being taken to infinity (i.e. incidence of a parallel beam), the photographic plate will be uniformly illuminated.

of the retina, are derived from the single ray SP, and the rays $P'M'$ and $Q'R'$, which focus at a different point O' on the retina, are derived from the ray SP'. Since the angles of refraction θ_1' and θ_2' (for these two sets of rays) will be different, the points O and O' will, in general, not have the same intensity.

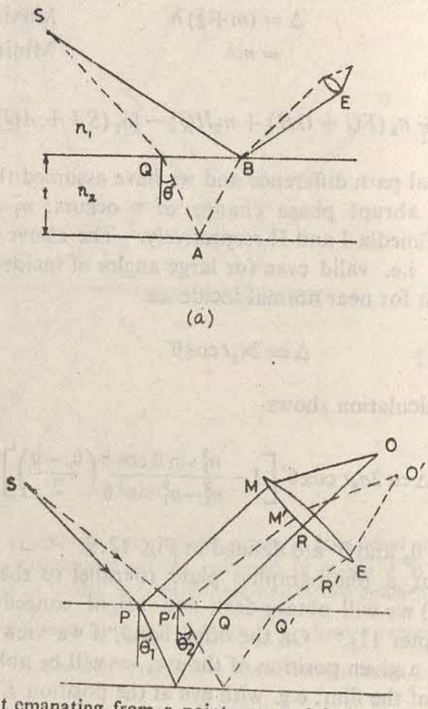


Fig. 12.10 Light emanating from a point source S is incident on a thin film: (a) if the film is viewed by the naked eye E then the point B will appear to be dark if the optical path $[\{n_1SQ+n_2\,(QA+AB)\}-n_1SB]$ is $m\lambda$ and bright if the optical path is $(m+\frac{1}{2})\lambda$. (b) If the eye is focussed for infinity then it receives parallel rays from different directions corresponding to different values of the angles of refraction (and hence different values of the optical path difference) correspond to different values of θ'.

We next consider the illumination by an extended source of light S (see Fig. 12.11). Such an extended source may be produced by illuminating a ground glass plate by a sodium lamp. Each point on the extended source will produce its own interference pattern on the photographic plate P, these will be displaced with respect to one another; consequently, no definite fringe pattern will appear on the photographic plate. However, if we view the film with our eye, rays from all points of the film will reach the eye. If the eye is focussed at infinity then parallel light coming in a particular direction reaching the eye would have originated from nearby points of the extended source and the intensity produced on the retina would depend on the value of $2nt\cos\theta'$ which is the same for all parallel rays like S_1Q, S_2Q', etc. (see Fig. 12.11). Rays emanating in a different direction (like S_1R, S_2R' etc.) would correspond to a different value of θ' and would focus at a different point on the retina. Since θ' is constant over the circumference of a cone (whose axis is normal to the film and whose vertex is at the eye), the eye will see dark and bright concentric rings, with the centre lying along the direction $\theta'=0$. Such fringes, produced by a film of uniform thickness, are known as Haidinger fringes. They are also known as fringes of equal inclination because the changes in the optical path are due to the changes in the direction of incidence and hence in the value of θ'. In Sec. 12.10 we will discuss the Michelson interferometer where such fringes are observed.

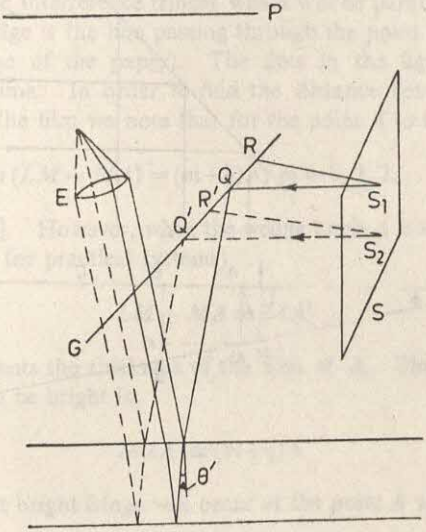


Fig. 12.11 Light emanating from an extended source illuminates a thin film. G represents the partially reflecting plate and P represents the photographic plate. The eye E is focussed at infinity.

12.7 Interference by a Film with Two Non-Parallel Reflecting Surfaces

Till now we have assumed the film to be of uniform thickness. We will now discuss the interference pattern produced by a film of varying thickness. Such a film may be produced by a wedge which consists of two non-parallel plane surfaces [see Fig. 12.12(a)].

We first consider a parallel beam of light incident normally on the upper surface of the film [see Fig. 12.12(a)]. In Fig. 12.12(b) the successive

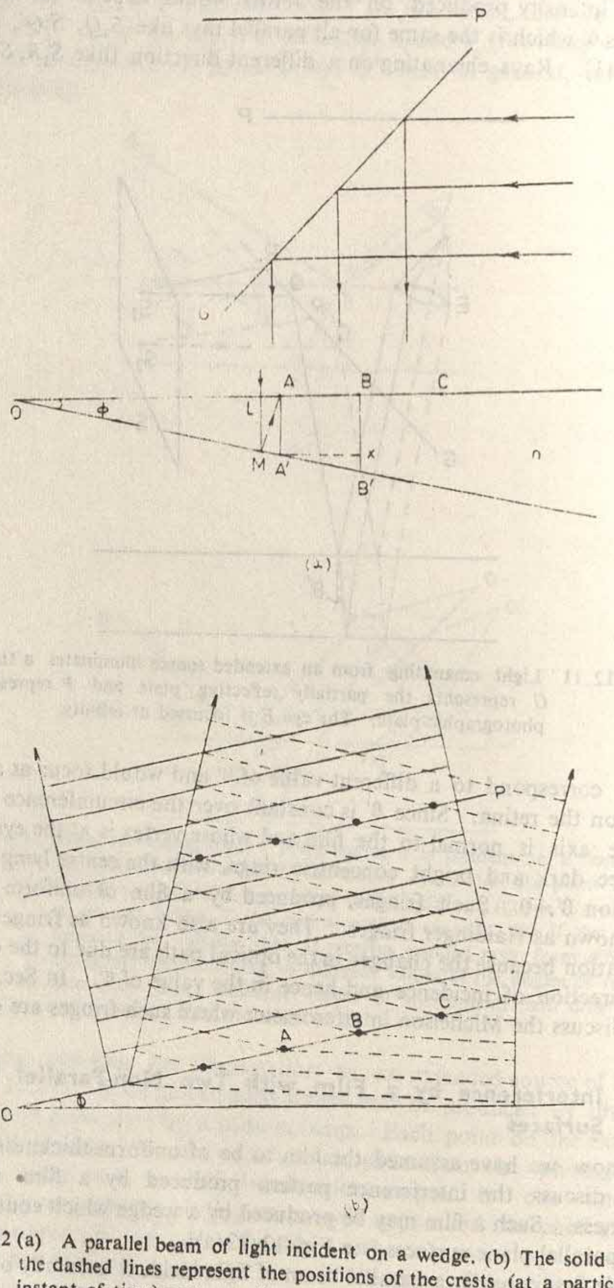


Fig. 12.12 (a) A parallel beam of light incident on a wedge. (b) The solid and the dashed lines represent the positions of the crests (at a particular instant of time) corresponding to the waves reflected from the upper surface and from the lower surface respectively. The maxima will correspond to the intersection of the solid and dashed lines. The fringes will be perpendicular to the plane of the paper.

positions of the crests (at a particular instant of time) reflected from the upper surface and from the lower surface of the film are shown by solid and dashed lines respectively. Obviously, a photographic plate P will record straight line interference fringes which will be parallel to the edge of the wedge (the edge is the line passing through the point O and perpendicular to the plane of the paper). The dots in the figure indicate the positions of maxima. In order to find the distance between two consecutive fringes on the film we note that for the point A to be bright*

$$n(LM+MA)=(m+\tfrac{1}{2})\lambda;\ m=0,1,2,\ldots \qquad (12.7\text{-}1)$$

[see Fig. 12.12(a)]. However, when the wedge angle φ is very small (which is indeed the case for practical systems)

$$LM+MA\approx 2AA'$$

where AA' represents the thickness of the film at A. Thus the condition for the point A to be bright is

$$2nAA'\approx(m+\tfrac{1}{2})\lambda \qquad (12.7\text{-}2)$$

Similarly, the next bright fringe will occur at the point B where

$$2nBB'\approx(m+\tfrac{3}{2})\lambda \qquad (12.7\text{-}3)$$

Thus

$$2n(BB'-AA')\approx\lambda$$

or

$$XB'\approx\lambda/2n \qquad (12.7\text{-}4)$$

But

$$XB'=(A'X)\tan\varphi$$

or

$$A'X=\beta\approx\frac{\lambda}{2n\varphi} \qquad (12.7\text{-}5)$$

where β represents the fringe width and we have assumed φ to be small. Such fringes are commonly referred to as *fringes of equal thickness.*

On the other hand, for a point source, the fringe pattern will be similar to the parallel film case; i.e. for near normal incidence, the pattern will be very nearly the same as produced by two sources S' and S'' (Fig. 12.13). (Notice that the point S'' is not vertically below S'; this is a consequence of the fact that the two surfaces of the film are not parallel). The intensity of an arbitrary point Q will be determined by the following equations:

$$\begin{aligned}[SA+n(AB+BC)+CQ]-[SD+DQ]&=(m+\tfrac{1}{2})\lambda \quad \text{Maxima}\\ &=m\lambda \quad \text{Minima}\end{aligned} \qquad (12.7\text{-}6)$$

---

*We are assuming here that the beam undergoes a sudden phase change of π when it gets reflected by the upper surface. The expression for the fringe width (Eq. 12.7-5) is, however, independent of this condition.

If we view the film with naked eye (say at the position E—see Fig. 12.13) then only a small portion of the film (around the point R) would be visible and the point R will be bright or dark as the optical path difference $[\{SN+n(NL+LR)\}-SR]$ is $(m+\frac{1}{2})\lambda$ or $m\lambda$ respectively. One can similarly discuss the case when the eye is focussed for infinity.

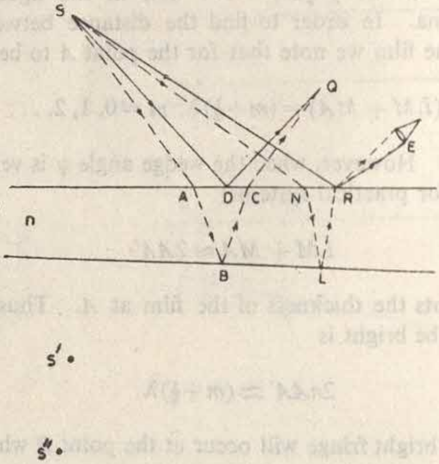


Fig. 12.13 Light from a point source illuminating a wedge. E represents the lens of the eye.

We next consider the illumination by an extended source S as shown in Fig. 12.14. Since the extended source can be assumed to consist of a

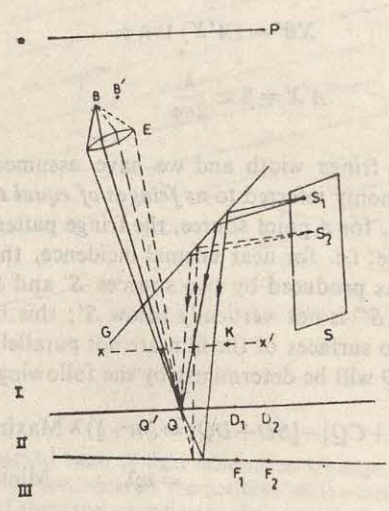


Fig. 12.14 Localized interference fringes produced by an extended source S. Fringes will be seen only when the eye is focussed on the upper surface of the film.

large number of independent point sources, each point source will produce its own pattern on a photographic plate P; consequently, no definite fringe pattern will be observed.* However, if we view the film with a camera (or with a naked eye) and if the camera is focussed on the upper surface of the film then a particular point on the film will appear dark or bright depending on the fact that whether $2nt$ is $m\lambda$ or $(m+\frac{1}{2})\lambda$ (see Fig. 12.14)—we are assuming near normal incidence. It may be seen in the figure that interference at the point Q may occur due to light coming from different points on the extended source, but if the incidence is near normal then the intensity at the point Q will be determined entirely by the thickness of the film at that place. Similarly, the intensity at the point Q' will be determined by the thickness of the film at Q'; however, the point Q' will be focussed at a different point B' on the retina of the eye. The fringes will be straight lines parallel to the edge of the film OO' (Fig. 12.15). It should be emphasized that all along we are assuming near normal incidence and the fact that the wedge angle is extremely small. These assumptions are indeed valid for practical systems.

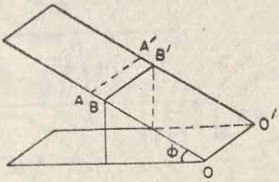


Fig. 12.15 The fringes formed by a wedge will be parallel to the edge OO'.

It is of interest to mention that if we focus the camera on a plane XX', which is slightly above the film, then no definite interference pattern will be observed. This follows from the fact that the light waves reaching the point K from S_2 undergo reflection at the points D_2 and F_2 and the light waves reaching K from S_1 undergo reflection at the points D_1 and F_1. Since the thickness of the film is not uniform, the waves reaching K from S_1 may produce brightness whereas, the waves reaching from S_2 may produce darkness. Thus, in order to view the fringes, one must focus the camera on the upper surface of the film, and in this sense, the fringes are said to be localized. It is left as an exercise for the reader to verify that if the camera is focussed for infinity, no definite interference pattern will be recorded.

Till now we have assumed the film to be 'thin' the question now arises as to how thin the film should be. In order to obtain an interference pattern there should be definite phase relationship between the waves reflected from the upper surface of the film and from the lower surface of the film. Thus the path difference $\Delta\ (= 2nt \cos \theta')$ should be small compared to the coherence length.* For example, if we are using the D_1 line of an ordinary sodium lamp ($\lambda = 5.890 \times 10^{-5}$ cm), the coherence length is of the order of 1 cm and for fringes to be visible Δ should be much less than 1 cm. It should be pointed out that there is no particular value of Δ for which the fringes disappear; but as the value of Δ increases, the contrast of the fringes becomes poorer. A laser beam has a very high coherence length and fringes can be visible even for path differences much greater than 1 m. On the other hand, if we use a white light source no fringes will be visible for $\Delta \gtrsim 2 \times 10^{-4}$ cm (see Sec. 11.9).

---

*There is, however, one exception to this, namely, when the extended source is taken to a very large distance, then the light rays reaching the plate G will be approximately parallel and an interference pattern (of low contrast) will be formed on the plate P. The same phenomenon will also occur if instead of moving the extended source we take the plate P far away from the wedge.

It should be pointed out that interference also occurs in region III (see Fig. 12.16) between the directly transmitted beam and the beam which comes out of the film after suffering two reflections, first from the lower surface and then from the upper surface of the film. However, the two amplitudes will be very different and the fringes will have very poor contrast (see Example 12.1).

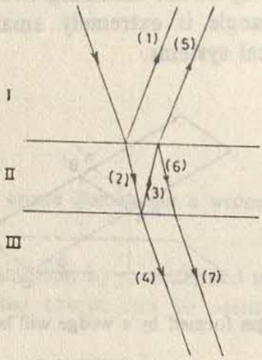


Fig. 12.16 In general, whereas the amplitudes of (1) and (5) are nearly the same, the amplitudes of (2) and (6) are quite different.

Example 12.1 Consider a film of refractive index 1.36 in air. Assuming near normal incidence ($\theta \approx 0$), show that whereas the amplitudes of the reflected rays (1) and (5) (Fig. 12.16) are nearly equal, the amplitudes of the transmitted rays (4) and (7) are quite different. (This is the reason why the fringes observed in transmission have very poor contrast).

Solution. Let the amplitude of the incident ray be a and let the amplitudes of the rays (1), (2), (3), ... be denoted by $a_1, a_2, \ldots$ etc. Using Eqs. (11.4-1a and 11.4-1b), we get

---

*Coherence length is defined in Sec. 14.1. If a source remains coherent for a time τ_c then the coherence length (L) will be about $c\tau_c$ where c is the speed of light in free space. Thus for $\tau_c \sim 10^{-10}$ sec, $L \sim 3$ cm.

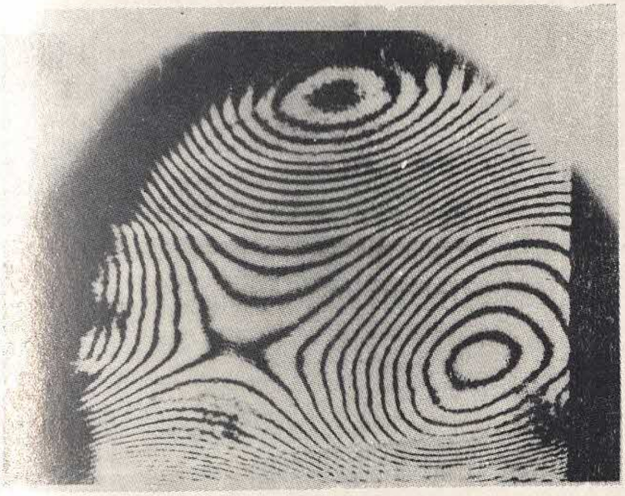


Plate 7 A typical fringe pattern produced by an air-film formed between two glass surfaces (which are not optically flat) and placed in contact with each other. Whenever the thickness of the air film is $m\lambda/2$, we obtain a dark fringe and when the thickness is $(m+\frac{1}{2})\ \lambda/2$, we obtain a bright fringe. Each fringe describes a locus of equal thickness of the film. (Photograph courtesy Dr. R. S. Sirohi.)

$$a_1 = \frac{1-n}{1+n}\,a = -\frac{0.36}{2.36}\,a \approx -0.153\,a$$

$$a_2 = \frac{2}{1+n}\,a = \frac{2}{2.36}\,a \approx 0.847\,a$$

$$a_3 = \frac{n-1}{n+1}\,a_2 = \frac{0.36}{2.36} \times 0.847\,a \approx 0.129\,a$$

$$a_5 = \frac{2n}{n+1}\,a_3 = \frac{2 \times 1.36}{2.36} \times 0.129a \approx 0.149a$$

$$a_4 = \frac{2n}{n+1}\,a_2 = \frac{2 \times 1.36}{2.36} \times 0.847a \approx 0.977a$$

$$a_7 = \frac{2n}{n+1}\,a_6 = \frac{2n}{n+1} \cdot \frac{n-1}{n+1}\,a_3 = \frac{2 \times 1.36 \times .36}{(2.36)^2}\,a_3 \approx 0.023a$$

We first note that the sign of a_5 is opposite to that of a_1 which is a consequence of the fact that a sudden phase change of π occurs when the ray gets reflected at the point B. Further the magnitude of a_5 is nearly equal to that of a_1. On the other hand $|a_7| \ll |a_4|$. This is the reason why the interference fringes formed in transmission have poor contrast.

12.8 Colours of Thin Films

We have seen in the previous section that if light from an extended monochromatic source (like a sodium lamp) is incident normally on a wedge, then equally spaced dark and bright fringes will be observed. The distance between two consecutive bright (or dark) fringes is determined by the wedge angle, the wavelength of light and by the refractive index of the film. If we use a polychromatic source (like an incandescent lamp) we will observe coloured fringes. Further, if instead of a wedge we have a film of an arbitrarily varying thickness we will again observe fringes, each fringe representing the locus of constant film thickness (see Plate 7). This is indeed what we see when sunlight falls on a soap bubble. It should be mentioned that if the optical path difference between the waves reflected from the upper surface of the film and from the lower surface of the film exceeds a few wavelengths, the interference pattern will be washed out due to the overlapping of interference patterns of many colours and no fringes will be seen (see Sec. 11.9). Thus, in order to see the fringes with white light, the film should not be more than few wavelengths thick.

12.9 Newton's Rings

If we place a plano convex lens on a plane glass surface, a thin film of air is formed between the curved surface of the lens (AOB) and the plane glass plate (POQ)—see Fig. 12.17. The thickness of the air film is zero at the point of contact O and increases as one moves away from the point of contact. If we allow monochromatic light (such as from a sodium lamp) to fall on the surface of the lens then the light reflected from the surface *AOB* interferes with the light reflected from the surface *POQ*. For near normal incidence (and considering points very close to the point of contact) the optical path difference between the two waves is very nearly equal to $2nt$, where n is the refractive index of the film and t the thickness of the film. Thus, whenever the thickness of the air film satisfies the condition

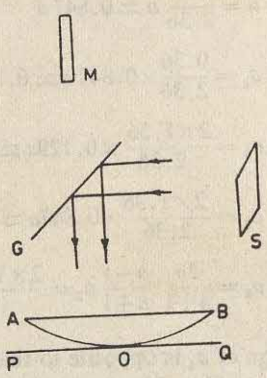


Fig. 12.17 An arrangement for observing Newton's rings. Light from an extended source *S* is allowed to fall on a thin film formed between the plano convex lens *AOB* and the plane glass plate *POQ*. *M* represents a travelling microscope.

$$2nt = (m + \tfrac{1}{2})\,\lambda; \; m = 0, 1, 2, \ldots \qquad (12.9\text{-}1)$$

we will have maxima. Similarly the condition

$$2nt = m\lambda \qquad (12.9\text{-}2)$$

will correspond to minima. Since the convex side of the lens is a spherical surface, the thickness of the air film will be constant over a circle (whose centre will be at O) and we will obtain concentric dark and bright rings. These rings are known as Newton's rings.* It should be pointed out that in order to observe the fringes, the microscope (or the eye) has to be focussed on the upper surface of the film (see the discussion in Sec. 12.7).

The radii of various rings can easily be calculated. As mentioned earlier, the thickness of the air film will be constant over a circle whose centre is at the point of contact O. Let the radius of the m^{th} dark ring be r_m and if t is the thickness of the air film where the m^{th} dark ring appears to be formed then

---

*Boyle and Hooke had independently observed the fringes earlier but Newton was the first to measure their radii and make an analysis. The proper explanation was given by Thomas Young.

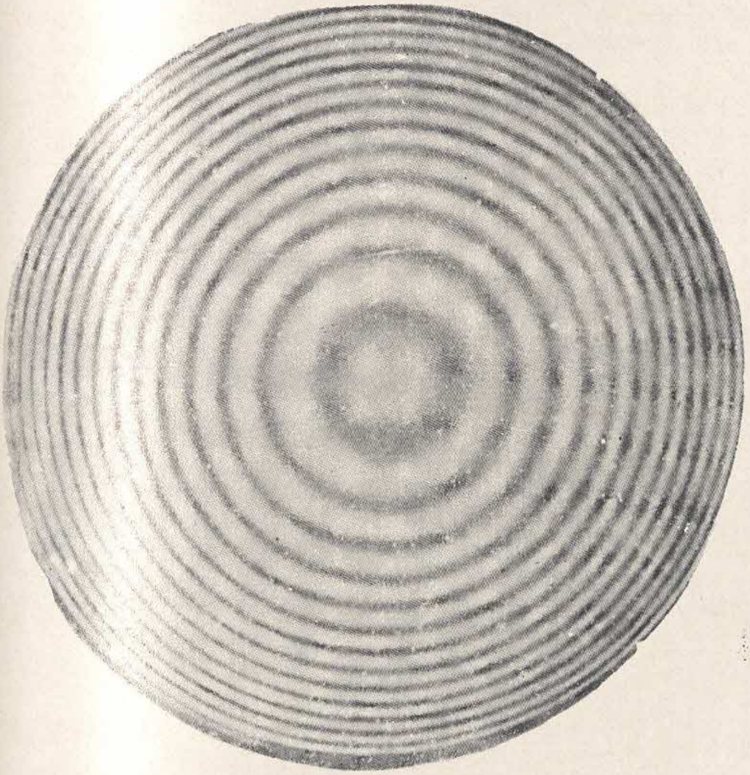


Plate 8 Newton's rings as observed in reflection. The rings observed with transmitted light are of much poorer contrast. (Photograph courtesy Mr. G. Bose.)

$$r_m^2 = t\,(2R-t) \tag{12.9-3}$$

where R represents the radius of curvature of the convex surface of the lens (see Fig. 12.18). Now $R \approx 100$ cm and $t \lesssim 10^{-3}$ cm, thus we may neglect t in comparison to $2R$ to obtain

$$r_m^2 \approx 2Rt$$

or

$$2t \approx \frac{r_m^2}{R} \tag{12.9-4}$$

Substituting this in Eq. (12.9-2) we get

$$r_m^2 \approx m\lambda R\,;\; m = 0,\, 1,\, 2, \tag{12.9-5}$$

which implies that the radii of the rings varies as square root of natural numbers. Thus the rings will become close to each other as the radius increases (see Plate 8). Between the two dark rings there will be a bright ring whose radius will be $\sqrt{m+\frac{1}{2}\lambda R}$.

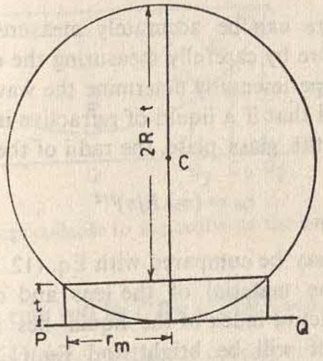


Fig. 12.18 r_m represents the radius of the mth dark ring; the thickness of the air film (where the mth dark ring is formed) is t.

Newton's rings can easily be observed in the laboratory by using an apparatus as shown in Fig. 12.17. Light from an extended source (emitting almost monochromatic light, like a sodium lamp) is allowed to fall on a glass plate which partially reflects the beam. This reflected beam falls on the plano convex lens-glass plate arrangement and Newton's rings can easily be observed by viewing directly or through a travelling microscope M. Actually, one really need not have a plano convex lens, the rings would be visible even when a biconvex lens is used.

Typically for $\lambda = 6 \times 10^{-5}$ cm and $R = 100$ cm

$$r_m = 0.0774\sqrt{m}\ \text{cm} \tag{12.9-6}$$

Thus the radii of the first, second and third dark rings would be approximately 0.0774 cm, 0.110 cm and 0.134 cm respectively. Notice that the spacing between the second and third dark ring is smaller than the spacing between the first and the second dark ring.

Equation (12.9-2) predicts that the central spot should be dark. Normally, with the presence of minute dust particles the point of contact is really not perfect and the central spot may not be perfectly dark. Thus while carrying out the experiment one should measure the radii of the m-th and the $(m+p)$-th ring ($p \approx 10$) and take the difference in the squares of the radii ($r_{m+p}^2 - r_m^2 = p\lambda R$) which is indeed independent of m. Usually, the diameter can be more accurately measured and in terms of the diameters the wavelength is given by the following expression:

$$\lambda = \frac{D_{m+p}^2 - D_m^2}{4pR} \tag{12.9-7}$$

The radius of curvature can be accurately measured with the help of a spherometer and therefore by carefully measuring the diameters of dark (or bright) rings one can experimentally determine the wavelength.

It may be mentioned that if a liquid of refractive index n is introduced between the lens and the glass plate, the radii of the dark rings would be given by

$$r_m = (m\lambda R/n)^{1/2} \tag{12 9-8}$$

The above equation may be compared with Eq. (12.9-5). Further, if the refractive indices of the material of the lens and of the glass plate are different and if the refractive index of the liquid lies in between the two values, the central spot will be bright and Eq. (12.9-8) would give the radii of the bright rings.

An important practical application of the principle involved in the Newton's rings experiment lies in the determination of the optical flatness of a glass plate. Consider a glass surface placed on another surface whose flatness is known. If a monochromatic light beam is allowed to fall on this combination and the reflected light is viewed by a microscope then, in general, dark and bright patches will be seen (Plate 7). The space between the two glass surfaces forms an air film of varying thickness and whenever this thickness becomes $m\lambda/2$, we see a dark spot and when this thickness becomes $(m+\frac{1}{2})\lambda/2$ we see a bright spot. Two consecutive dark fringes will be separated by the air film whose thickness will differ by $\lambda/4$. Consequently, by measuring the distance between consecutive dark and bright fringes one can calculate the optical flatness of a glass plate.

Example 12.2 Consider the formation of Newton's rings by monochromatic light of $\lambda = 6.4 \times 10^{-5}$ cm. Assume the point of contact to be perfect. Now slowly raise the lens vertically above the plate. As the lens moves gradually away from the plate discuss the ring pattern as seen through the microscope. Assume the radius of the convex surface to be 100 cm.

Solution. Since the point of contact is perfect the central spot will be dark, the first dark ring will form at P where $PA=\lambda/2$ and the radius of this ring, OA, will be $\sqrt{\lambda R}$ ($=0.080$ cm)—see Fig. 12.19(a). Similarly, the radius of the second dark ring will be $OB=\sqrt{2\lambda R}$ ($=0.113$ cm). If we now raise the lens by $\frac{\lambda}{4}$ ($=1.6\times10^{-5}$ cm) then $2t$ corresponding to the central spot would be $\lambda/2$ and instead of the dark spot at the centre we will now have a bright spot. The radii of the first and the second dark rings will be

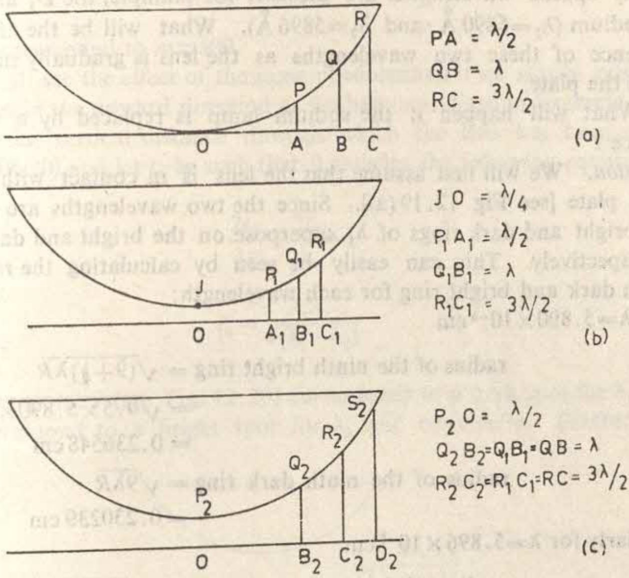


Fig. 12.19 The rings collapse to the centre as the lens is moved away from the plate.

$$OA_1=\sqrt{\tfrac{1}{2}\lambda R}=0.0566\text{ cm}$$

and

$$OB_1=\sqrt{\tfrac{3}{2}\lambda R}=0.098\text{ cm}$$

respectively [see Fig. 12.19(b)]. If the lens is further moved by $\lambda/4$ (see Fig. 12.19(c)], then the first dark ring collapses to the centre and the central spot will be dark. The ring which was originally at Q now shifts to Q_2; similarly the ring at R [Fig. 12.19(a)] collapses to R_2 [Fig. 12.19(c)].

Thus, as the lens is moved upward the rings collapse to the centre. Hence if we can measure the distance by which the lens is moved upward and also count the number of dark spots that have collapsed to the centre, we can determine the wavelength. For example, in the present case, if the lens is moved by 6.4×10^{-3} cm, 200 rings will collapse to the centre. If one carries out this experiment it will be observed that the 200th dark ring will slowly converge to the centre and when the lens has moved exactly by 6.4×10^{-3} cm it has exactly come to the centre.

Example 12.3 Consider the formation of Newton's rings when two closely spaced wavelengths are present; for example, the D_1 and D_2 lines of sodium ($\lambda_1=5890$ Å and $\lambda_2=5896$ Å). What will be the effect of the presence of these two wavelengths as the lens is gradually moved away from the plate.

What will happen if the sodium lamp is replaced by a white light source?

Solution. We will first assume that the lens is in contact with the plane glass plate [see Fig. 12.19 (a)]. Since the two wavelengths are very close, the bright and dark rings of λ_1 superpose on the bright and dark rings of λ_2 respectively. This can easily be seen by calculating the radii of the ninth dark and bright ring for each wavelength:

For $\lambda=5.890\times10^{-5}$ cm

$$\begin{aligned}\text{radius of the ninth bright ring} &= \sqrt{(9+\tfrac{1}{2})\lambda R}\\ &= \sqrt{9.5\times5.890\times10^{-5}\times100}\\ &= 0.236548\text{ cm}\\ \text{radius of the ninth dark ring} &= \sqrt{9\lambda R}\\ &= 0.230239\text{ cm}\end{aligned}$$

Similarly for $\lambda=5.896\times10^{-5}$ cm

$$\begin{aligned}\text{radius of the ninth bright ring} &= \sqrt{9.5\times5.896\times10^{-5}\times100}\\ &= 0.23666\text{ cm}\end{aligned}$$

and

$$\begin{aligned}\text{radius of the ninth dark ring} &= \sqrt{9\times5.896\times10^{-3}}\\ &= 0.230356\text{ cm}\end{aligned}$$

Thus the rings almost exactly superpose on each other. However, for large values of m, the two ring patterns may produce uniform illumination. To be more specific, when the air film thickness t is such that

$$2t = m\lambda_1 = (m+\tfrac{1}{2})\lambda_2$$

or

$$\frac{2t}{\lambda_2}-\frac{2t}{\lambda_1}=\frac{1}{2} \qquad (12.9\text{-}9)$$

then around that point the fringe system will completely disappear; i.e. the bright ring for the wavelength λ_1, will fall on the dark ring for the wavelength λ_2 and conversely. Thus the contrast will be zero and no fringe pattern will be visible. Rewriting Eq. (12.9-9) we get

$$2t\,\frac{\lambda_1-\lambda_2}{\lambda_1\lambda_2}=\frac{1}{2}$$

or

$$2t = \tfrac{1}{2}\frac{\lambda_1\lambda_2}{\Delta\lambda} \approx \tfrac{1}{2}\frac{(5.893\times10^{-5})^2}{6\times10^{-8}}$$

$$\approx 3\times10^{-2}\,\text{cm}$$

This will correspond to $m \approx 500$.

We shall see the effect of the same phenomenon if we slowly raise the convex lens in the upward direction as we had considered in Example 12.1. Let t_0 be the vertical distance through which the lens has been raised (see Fig. 12.20) and let t_0 be such that it satisfies the following equation

$$\frac{2t_0}{\lambda_2} - \frac{2t_0}{\lambda_1} = \tfrac{1}{2}$$

or

$$t_0 = \frac{\lambda_1\lambda_2}{4(\lambda_1-\lambda_2)}$$

Thus, if the point J (see Fig. 12.20) corresponds to a dark spot for λ_1 then it will correspond to a bright spot for λ_2 and conversely. Further, the nearby dark rings for λ_1, will almost fall at the same place as the bright rings for λ_2 and the interference pattern will be washed out. Thus viewing from a microscope we will not be able to see any ring pattern. Now, if the lens is further moved upwards by a distance t_0 then we will have

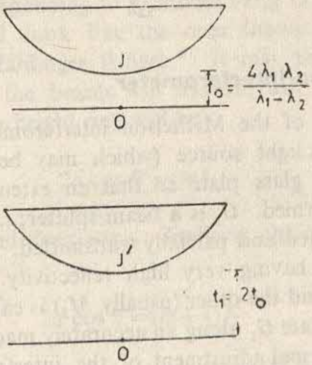


Fig. 12.20 In the Newton's ring experiment if the light consists of two closely spaced wavelengths λ_1 and λ_2 (like the D_1 and D_2 lines of sodium) then, if the lens is separated by a distance $t_0\left(=\frac{4\lambda_1\lambda_2}{\lambda_1-\lambda_2}\right)$ interference fringes will be washed out. The fringes will reappear when the distance is $2t_0$.

$$\frac{2t_1}{\lambda_2} - \frac{2t_1}{\lambda_1} = 1 \qquad (12.9\text{-}10)$$

where $t_1=2t_0$. Consequently, if the point J' corresponds to a dark spot for λ_1 then it will also correspond to a dark spot for λ_2. The fringe pattern will reappear but now with a slightly weaker contrast (why?).

In this way if we continue to move the lens upwards the fringe system will reappear everytime the lens is moved up by a distance $2t_0\left(\approx\frac{1}{2}\,\frac{\lambda_1\lambda_2}{\Delta\lambda}\right)$. This principle is used in a Michelson interferometer to measure the small wavelength difference $\Delta\lambda$, between two closely spaced lines (like the D_1 and D_2 lines of sodium).

It should be pointed out that for complete disappearance of the fringe pattern the intensities of the two lines λ_1 and λ_2 should be the same.

Another corollary of the above experiment consists in finding the change in the interference pattern (as we move up the convex lens) when we consider a single line of wavelength λ, but which has a width of $\Delta\lambda$. Thus we should assume all wavelengths between λ and $\lambda+\Delta\lambda$ to exist. By finding the approximate height at which the fringes disappear one can calculate $\Delta\lambda$. The coherence length (L) is related to $\Delta\lambda$ through the following relation (see Sec. 14.2):

$$L\sim\frac{\lambda^2}{\Delta\lambda} \qquad (12.9\text{-}11)$$

12.10 The Michelson Interferometer

A schematic diagram of the Michelson interferometer is shown in Fig. 12.21. S represents a light source (which may be a sodium lamp) and L represents a ground glass plate so that an extended source of almost uniform intensity is formed. G_1 is a beam splitter; i.e., a beam incident on G_1 gets partially reflected and partially transmitted. M_1 and M_2 are good quality plane mirrors having very high reflectivity. One of the mirrors (usually M_2) is fixed and the other (usually M_1) is capable of moving away or towards the glass plate G_1 along an accurately machined track by means of a screw. In the normal adjustment of the interferometer, the mirrors M_1 and M_2 are perpendicular to each other and G_1 is at 45° to the mirror.

Waves emanating from a point P get partially reflected and partially transmitted by the beam splitter G_1, and the two resulting beams are made to interfere in the following manner: The reflected wave [shown as (1) in Fig. 12.21] undergoes a further reflection at M_1 and this reflected wave gets (partially) transmitted through G_1; this is shown as (5) in the figure. The transmitted wave [shown as (2) in Fig. 12.21] gets reflected by M_2 which gets (partially) reflected by G_1 and results in the wave shown as (6) in the figure. Waves (5) and (6) interfere in a manner exactly similar to that shown in Fig. 12.11. This can easily be seen from the fact that if x_1 and x_2 are the distances of the mirrors M_1 and M_2 from the plate G_1, then to the eye the waves emanating from the point P will appear to get reflected by two parallel mirrors [M_1 and M_2'—see Fig. 12.21] separated

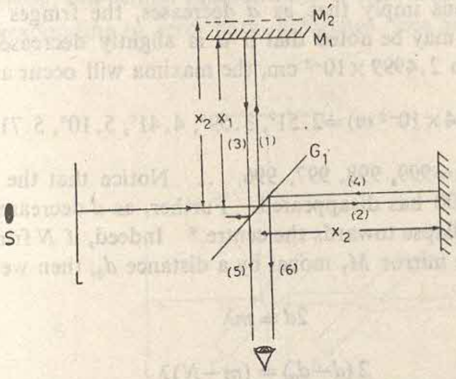


Fig. 12.21 Schematic of the Michelson interferometer.

by a distance $(x_1 \sim x_2)$. As discussed in Sec. 12.6, if we use an extended source, then no definite interference pattern will be obtained on a photographic plate placed at the position of the eye. Instead, if we have a camera focussed for infinity then on the focal plane we will obtain circular fringes, each circle corresponding to a definite value of θ' (see Fig. 12.11); the circular fringes will look like the ones shown in Plate 9a. These fringes are known as Haidinger fringes.* It may be mentioned that the extra path that one of the beams will traverse will be $2(x_1 \sim x_2)$ and as such, the condition for a bright ring will be

$$2d \cos \theta' = m\lambda \qquad (12.10\text{-}1)$$

where m is an integer and $d = x_1 \sim x_2$. Similarly, the condition for a dark ring would be

$$2d \cos \theta' = (m + \tfrac{1}{2})\lambda \qquad (12.10\text{-}2)$$

For example, for $\lambda = 5 \times 10^{-5}$ cm if $d = 2.5 \times 10^{-2}$ cm, the angles at which the bright rings will occur will be

$$\theta' = \cos^{-1}(0.001\, m) = 0°,\ 2.56°,\ 3.62°,\ 4.44°,\ 5.13°,\ 5.73°,\ 6.28°, \ldots$$

where the angles correspond to $m = 1000, 999, 998, 997, 996, 995, \ldots$ On the other hand, for $d = 2.5 \times 10^{-3}$ cm, the corresponding angles will be

$$\theta' = \cos^{-1}(0.01\, m) = 0°,\ 8.11°,\ 11.48°,\ 14.07°,\ 16.26°,\ 18.19°,\ 19.95°, \ldots$$

where the angles correspond to $m = 100, 99, 98, 97, 96, 95, \ldots$

---

*It is left as an exercise for the reader to discuss the interference pattern when the two mirrors M_1 and M_2 are not exactly perpendicular to each other (see Fig. 12.14).

The above equations imply that as d decreases, the fringes become less closely spaced. It may be noted that if d is slightly decreased, say from 2.5000×10^{-2} cm to 2.4999×10^{-2} cm, the maxima will occur at

$$\theta' = \cos^{-1}(1.00004 \times 10^{-3}\, m) = 2.51°, 3.59°, 4.41°, 5.10°, 5.71°, 6.26°, \ldots$$

corresponding to $m=999, 998, 997, 996, \ldots$ Notice that the ring corresponding to $m=1000$ has disappeared. Further, as d decreases, the fringe pattern tends to collapse towards the centre.* Indeed, if N fringes collapse to the centre as the mirror M_1 moves by a distance d_0, then we must have

$$2d = m\lambda \tag{12.10-3}$$

$$2(d-d_0) = (m-N)\lambda \tag{12.10-4}$$

where we have put $\theta' = 0$ because we are looking at the central fringe. Thus,

$$\lambda = \frac{2d_0}{N} \tag{12.10-5}$$

This provides us with a method for the measurement of the wavelength. For example, in a typical experiment, if one finds 1000 fringes collapse to the centre as the mirror is moved through a distance of 2.90×10^{-2} cm, then

$$\lambda = 5800\ \text{Å}$$

The above method was used by Michelson for the standardization of the meter. He had found that the red cadmium line ($\lambda = 6438.4696$ Å) is one of the ideal monochromatic sources and as such this wavelength was used as a reference for the standardization of the meter. In fact he defined the meter by the following relation:

1 meter = 1553164.13 red cadmium wavelengths

It may be noted that the accuracy is almost one part in 10^9.

In an actual Michelson's interferometer, the beam splitter G_1 consists of a plate (which may be about 1/2 cm thick), the back surface of which is partially silvered and the reflections occur at the back surface as shown in Fig. 12.22. It is immediately obvious that the beam (5) traverses the glass plate thrice and in order to compensate for this additional path, one introduces a 'compensating plate' G_2 which is exactly of the same thickness as G_1. It may be mentioned that the compensating plate is not really necessary for a monochromatic source because the additional path $2t$ introduced by G_1 can be compensated by moving the mirror M_1 by a distance nt where n is the refractive index of the material of the glass G_1. However, for a white light source, one would observe only a few coloured fringes around the point corresponding to zero path difference (see Sec. 11.9) and, since

---

*Conversely, if d is increased, the fringe pattern will expapd,

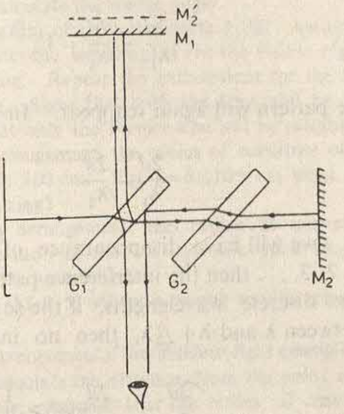


Fig. 12.22 In an actual interferometer there is also a compensating plate G_2.

the refractive index depends on the wavelength, it is necessary to have the compensating plate to observe the fringe pattern. For example, for $\lambda=6560$ Å and 4861 Å, the refractive index of crown glass is 1.5244 and 1.5330 respectively. As such, 0.5 cm thick crown glass plate will contain about 11619 of 6560 Å wavelengths and about 15768 of 4861 Å wavelengths. Thus it is obvious that if we have a continuous range of wavelengths from 4861 Å to 6560 Å, we should have a compensating plate to observe the fringe pattern.

Michelson interferometer can also be used in the measurement of two closely spaced wavelengths. Let us assume that we have a sodium lamp which emits predominantly two closely spaced wavelengths 5890 Å and 5896 Å. The interferometer is first set corresponding to the zero path difference.* Near $d=0$, both the fringe patterns will overlap. If the mirror M_1 is moved away (or towards) the plate G_1 through a distance d, then the maxima corresponding to the wavelength λ_1 will not, in general, occur at the same angle as λ_2. Indeed, if the distance d is such that

$$\frac{2d}{\lambda_1}-\frac{2d}{\lambda_2}=\frac{1}{2} \qquad (12.10\text{-}6)$$

---

*The zero path difference is easily obtained by using white light where only a few coloured fringes, around $d=0$, will be visible.

and if $2d\cos\theta' = m\lambda_1$, then $2d\cos\theta' = (m+\frac{1}{2})\,\lambda_2$. Thus, the maxima of λ_1 will fall on the minima of λ_2 and conversely, and the fringe system will disappear. It can easily be seen that if

$$\frac{2d}{\lambda_1} - \frac{2d}{\lambda_2} = 1 \tag{12.10-7}$$

then interference pattern will again reappear. In general, if

$$\frac{2d}{\lambda_1} - \frac{2d}{\lambda_2}$$

is 1/2, 3/2, 5/2 . . . we will have disappearance of the fringe pattern and if it is equal to 1, 2, 3, . . . then the interference pattern will appear.

Instead of two discrete wavelengths, if the source consists of all wavelengths, lying between λ and $\lambda+\triangle\lambda$, then no interference pattern will be observed if

$$\frac{2d}{\lambda} - \frac{2d}{\lambda + \frac{\Delta\lambda}{2}} \gtrsim \frac{1}{2}$$

or

$$2d \geq \frac{\lambda^2}{\Delta\lambda} \tag{12.10-8}$$

In this case the fringes will not reappear because we have a continuous range of wavelengths rather than two discrete wavelengths (see Sec. 14.2).

Example 12.4 For a sodium lamp, the distance traversed by the mirror between two successive disappearances is 0.289mm. Calculate the difference in the wavelengths of the D_1 and the D_2 lines. Assume $\lambda \simeq 5890$ Å.

Solution. When the mirror moves through a distance 0.289 mm. the additional path introduced is 0.578 mm. Thus

$$\frac{0.578}{\lambda} - \frac{0.578}{\lambda+\Delta\lambda} = 1$$

or

$$\Delta\lambda \simeq \frac{\lambda^2}{0.578} = \frac{(5890\times10^{-7})^2}{0.578}\ \text{mm}$$

$$\simeq 6\ \text{Å}$$

PROBLEMS

1. A glass plate (of refractive index 1.6) is in contact with another glass plate (of refractive index 1.8) along a line such that a wedge of 0.5° is formed. Light of wavelength 5000 Å is incident vertically on the wedge and the film is viewed from the top. Calculate the fringe spacing. The whole apparatus is immersed in an oil of refractive index 1.7. What will be the qualitative difference in the fringe pattern and what will be the new fringe width?

2. Two plane glass plates are placed on top of one another and on one side a cardboard is introduced to form a thin wedge of air. Assuming that a beam of wavelength 6000 Å is incident normally, and that there are 100 interference fringes per centimeter, calculate the wedge angle.
3. Consider a nonreflecting film of refractive index 1.38. Assume that its thickness is 9×10^{-6} cm. Calculate the wavelengths (in the visible region) for which the film will be non-reflecting. Repeat the calculations for the thickness of the air film to be 45×10^{-6} cm. Show that both the films will be non-reflecting for a particular wavelength but only the former one will be suitable. Why?
4. In the Newton's rings arrangement, the radius of curvature of the curved side of the plano convex lens is 100 cm. For $\lambda=6\times10^{-5}$ cm what will be the radii of the 9th and 10th bright ring?
5. In the Newton's rings arrangement, the radius of curvature of the curved surface is 50 cm. The radii of the 9th and 16th dark rings are 0.18 cm and 0.2235 cm. Calculate the wavelength.
   (Hint: The use of Eq. (12.9–5) will give wrong results, why?)
   (Ans. ≈5000 Å)
6. In the Newton's rings arrangement if the incident light consists of two wavelengths (4000 Å and 4002 Å) calculate the distance (from the point of contact) at which the rings will disappear. Assume that the radius of curvature of the curved surface is 80 cm.
7. In the above problem if the lens is slowly moved upward, calculate the height of the lens at which the fringe system (around the centre) will disappear.
   (Ans. 0.2 mm).
8. An equiconvex lens is placed on another equiconvex lens. The radii of curvatures of the two surfaces of the upper lens are 50 cm and those of the lower lens are 18 cm. The waves reflected from the upper and lower surface of the air film (formed between the two lenses) interfere to produce Newton's rings. Calculate the radii of the dark rings. Assume $\lambda=6000$ Å.
9. In the Michelson interferometer arrangement, if one of the mirrors is moved by a distance 0.08 mm, 250 fringes cross the field of view. Calculate the wavelength.
   (Ans. 6400 Å)
10. The Michelson's interferometei experiment is performed with a source which consists of two wavelengths 4882 Å and 4886 Å. Through what distance does the mirror have to be moved between two positions of the disappearance of the fringes?
11. In the Michelson interferometer experiment, calculate the various values of θ' (corresponding to bright rings) for $d=5\times10^{-3}$ cm. Show that if d is decreased to 4.997×10^{-3} cm the fringe corresponding to $m=200$ disappears. What will be the corresponding values of θ'? Assume $\lambda=5\times10^{-5}$ cm.

REFERENCES AND SUGGESTED READING

See at the end of Chapter 13

13 Multiple Beam Interferometry

> *There was no 'one, two, three and away', but they began running when they liked and left off when they liked, so that it was not easy to know when the race was over.*
>
> LEWIS CARROLL

13.1 Introduction

In the last two chapters, we have been discussing interference between two beams which are derived from a single beam either by division of wavefront or by division of amplitude. In this chapter, we will discuss the interference involving many beams which are derived from a single beam by multiple reflections (division of amplitude). Thus, for example, if a plane wave falls on a plane parallel glass plate then the beam would undergo multiple reflections at the two surfaces and a large number of beams of successively diminishing amplitude will emerge on both sides of the plate. These beams (on either side) interfere to produce an interference pattern at infinity. We will show that the fringes so formed are much sharper than those formed by two beam interference and, therefore, the interferometers involving multiple beam interference have a very high resolving power and hence find applications in high resolution spectroscopy.

13.2 Multiple Reflections from a Plane Parallel Film

Let us consider a plane parallel film of thickness h and of refractive index n, surrounded by a medium of refractive index n', illuminated by a plane wave at a finite angle (see Fig. 13.1). The wave will suffer multiple reflections at the two parallel surfaces of the film.

Let A_0 be the complex amplitude of the wave incident at the film at an angle θ_i. The wave undergoes multiple reflections at the two surfaces as shown in Fig. 13.1. If δ represents the phase difference between two

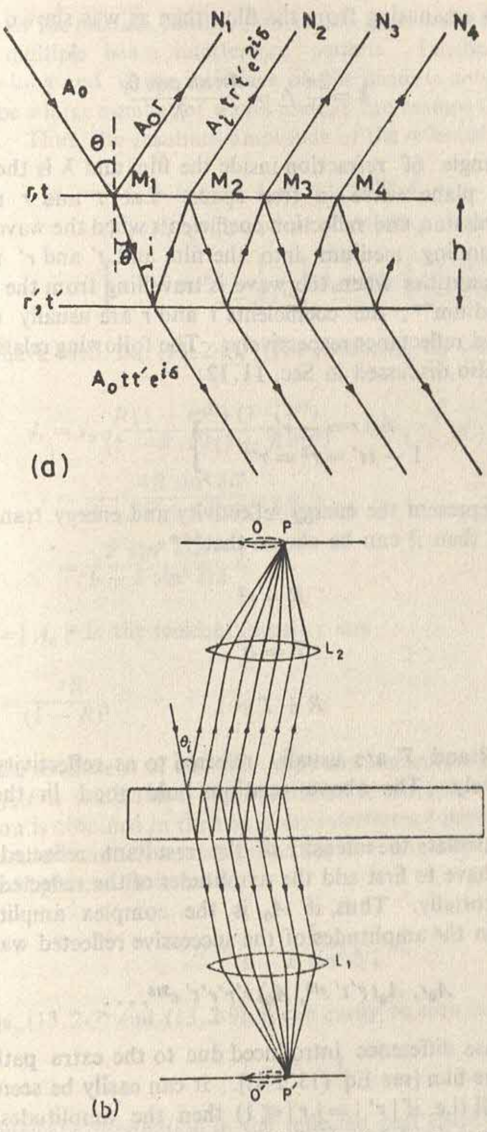


Fig. 13.1 (a) Reflection and transmission of a beam of amplitude A_0 incident at an angle θ_i on a film of refractive index n and thickness h.

(b) Any ray parallel to AB will focus at the same point P. If the ray AB is rotated about the normal at B, then the point P will rotate on the circumference of a circle centred at the point 0; this circle will be bright or dark depending on the value of θ_i. Rays incident at different angles will focus at different distances from the point 0 and one will obtain concentric bright and dark rings for an extended source.

successive waves emanating from the film, then as was shown in Sec. 12.3, we have*

$$\delta = \frac{2\pi}{\lambda} \Delta = \frac{4\pi nh \cos \theta_r}{\lambda} \tag{13.2-1}$$

where θ_r is the angle of refraction inside the film and λ is the wavelength of the incident plane wave in free space. Let t and r represent the amplitude transmission and reflection coefficients when the wave is travelling from the surrounding medium into the film and t' and r' represent the corresponding quantities when the wave is travelling from the film into the surrounding medium**; the coefficients t and r are usually called as the transmittance and reflectance respectively. The following relations (Stokes' relations) were also discussed in Sec. 11.12:

$$\left.\begin{array}{c} r = -r' \\ 1 - tt' = r^2 = r'^2 \end{array}\right\} \tag{13.2-2}$$

If R and T represent the energy reflectivity and energy transmittivity of the film surfaces, then it can be shown that***

$$R = r^2 \tag{13.2-3}$$

and

$$T = tt' \tag{13.2-4}$$

which gives

$$R + T = 1 \tag{13.2-5}$$

The coefficients R and T are usually referred to as reflectivity and transmittivity respectively. The above relations hold good in the absence of absorption only.

In order to calculate the intensity of the resultant reflected and transmitted light, we have to first add the amplitudes of the reflected and transmitted waves vectorially. Thus, if A_0 is the complex amplitude of the incident light, then the amplitudes of the successive reflected waves will be

$$A_0 r,\ A_0 t r' t' e^{i\delta},\ A_0 t r' r' r' t' e^{2i\delta}, \ldots$$

where δ is the phase difference introduced due to the extra path traversed by the beam in the film [see Eq. (13.2-1)]. It can easily be seen that if the reflectivity is small (i.e. if $|r'| = |r| \ll 1$) then the amplitudes of the first two waves are approximately equal and the other waves have negligible amplitudes. Thus, we will essentially have a two beam interference as discussed in Chapter 12. On the other hand, if the reflectivity is large, then the amplitudes of the second, third ... waves are approximately equal and one obtains a multiple beam interference pattern. Further, if the plate is sufficiently long and if the incidence of the beam is not too oblique, then there will be a large number of waves and we can assume them to be infinite in number. Thus, the resultant amplitude of the reflected wave would be

---

*We are neglecting here any possible phase shifts which might arise due to reflections at the surfaces. If the surfaces are coated with a metal, there would arise additional phase shifts on reflection at the two surfaces.

**Explicit expressions for t and r will be evaluated in Chapter 20.

***The energy reflection and transmission coefficients are discussed in Chapter 20.

$$\begin{aligned} A_r &= r A_0 + tr't' e^{i\delta} A_0 + tt' r'^3 e^{2i\delta} A_0 + \dots \\ &= [r + tt'r' e^{i\delta} (1 + r'^2 e^{i\delta} + r'^4 e^{2i\delta} + \dots)] A_0 \\ &= \left[r - Tr e^{i\delta} \frac{1}{1 - R e^{i\delta}} \right] A_0 = A_0 r \left[\frac{1 - e^{i\delta}}{1 - R e^{i\delta}} \right] \end{aligned} \tag{13.2-6}$$

where we have used Eq. (13.2-5). The intensity, I_r, will be proportional to $|A_r|^2$:

$$\begin{aligned} I_r &= I_0 \frac{R(1 - e^{-i\delta})(1 - e^{i\delta})}{(1 - R e^{i\delta})(1 - R e^{-i\delta})} = I_0 R \frac{(2 - 2\cos\delta)}{1 + R^2 - 2R\cos\delta} \\ &= \frac{4R \sin^2 \delta/2}{(1 - R)^2 + 4R \sin^2 \delta/2} I_0 \\ &= \frac{F \sin^2 \delta/2}{1 + F \sin^2 \delta/2} I_0 \end{aligned} \tag{13.2-7}$$

where $I_0 = |A_0|^2$ is the incident intensity and

$$F = \frac{4R}{(1 - R)^2} \tag{13.2-8}$$

is called the coefficient of Finesse. One can immediately see that for small reflectivity, F is small and I_r is proportional to $\sin^2 \delta/2$. The same intensity distribution is obtained in the two beam interference pattern (see Fig. 11.8).

In a similar manner, an expression for the intensity distribution of the transmitted light can be derived:

$$I_t = \frac{1}{1 + F \sin^2 \delta/2} I_0 \tag{13.2-9}$$

From Eqs. (13.2-7) and (13.2-9), it can easily be seen that

$$I_r + I_t = I_0 \tag{13.2-10}$$

i.e., the intensity distribution in the reflected part and the transmitted part are complimentary to each other.

In the above analysis absorption has been neglected. However, if absorption is also included in the analysis, the intensity of the transmitted beam becomes

$$I_t = \left(1 - \frac{A}{1 - R}\right)^2 \frac{I_0}{1 + F \sin^2 \delta/2} \tag{13.2-11}$$

where A is the absorption coefficient of the reflecting surface such that

$$R+T+A=1 \qquad (13.2\text{-}12)$$

We now consider the case when the film is illuminated by plane waves, having nearly equal amplitudes and incident at various angles. Corresponding to each wave there will be a set of parallel waves in the reflected part as well as in the transmitted part. The transmitted light is collected by a lens L_1 which focusses each set of parallel waves into a point in its focal plane. Now, according to Eq. (13.2-9), the maximum of intensity in the transmitted light will correspond to

$$\delta = 2m\pi \qquad (13.2\text{-}13)$$

where m is an integer. For a particular film (i.e., for given n and h), the maxima will correspond to those angles θ_r for which

$$\cos\theta_r = \frac{m\lambda}{2nh}\,;\ m=1, 2, 3, \ldots \qquad (13.2\text{-}14)$$

Thus the fringes in the focal plane of L_1 will be the loci of points corresponding to constant θ_r (and hence θ_i) and are known as fringes of equal inclination. The minima will correspond to $\delta=(2m+1)\,\pi$; $m=0, 1, 2, \ldots$ Similarly, if the reflected part is collected by a lens L_2, the maxima of the fringe system will correspond to [see Eq. (13.2-7)];

$$\delta=(2m+1)\pi;\ m=0, 1, 2, \ldots \qquad (13.2\text{-}15)$$

and minima will correspond to

$$\delta=2m\pi;\ m=1, 2, 3, \ldots \qquad (13.2\text{-}16)$$

Thus, the two fringe systems are complementary to each other. This is also evident from Eq. (13.2-10). Thus on illuminating the film with an extended source, one observes in the focal planes of L_1 and L_2 a system of fringes. The fringes would be circular with their centre at the point corresponding to the wavefront emerging normally from the film [see Fig. 13.1(b)].

The sharpness of fringes (determined by the width of the fringes) depends on the value of the parameter F and, hence, on the reflectivity R. In Fig. 13.2, we have plotted the transmitted intensity I_t/I_0 as a function of δ for different values of F. It can be seen from the figure that the fringes become sharper as the value of F increases. If we define the half width of a fringe as the full width of the fringe corresponding to half of the maximum intensity, then (see Problem 13.2)

$$\text{half width} = \frac{4}{\sqrt{F}} = \frac{2(1-R)}{\sqrt{R}} \qquad (13.2\text{-}17)$$

Thus as F increases, the half width decreases and the fringes become sharper. For large values of F the transmitted fringe system consists of very narrow bright fringes with an almost completely dark background whereas the reflected pattern will have dark fringes with a bright background.

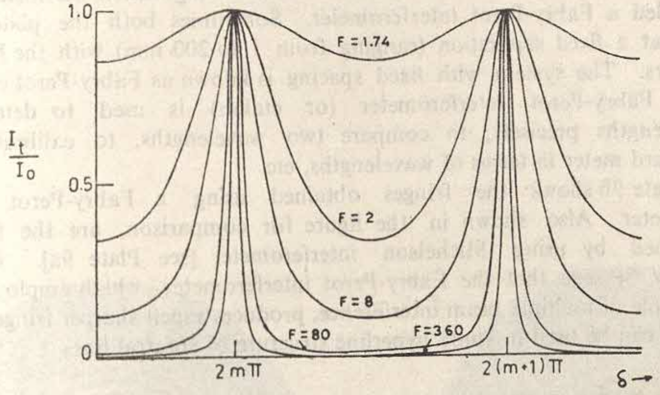


Fig. 13.2 The transmitted intensity as a function of δ for different values of F.

13.3 Fabry-Perot Interferometer

In this section, we will discuss the Fabry-Perot interferometer which is based on the principle of multiple beam interferometry discussed in the last section. The interferometer (as shown in Fig. 13.3) consists of two plane glass (or quartz) plates which are coated on one side with a partially reflecting metallic film* (of aluminum or silver) of about 80% reflectivity. These two plates are kept in such a way that they enclose a plane parallel slab of air between their coated surfaces. In order to have sharp fringes, the coated surfaces should be parallel to a high degree of accuracy and the surfaces themselves should be flat upto a very small fraction of wavelength (flatness of the order of $\lambda/20$ to $\lambda/100$ is required, where λ is the wavelength of the light to be used). The two uncoated surfaces of

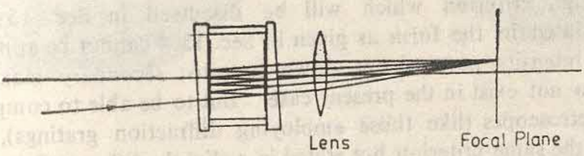


Fig. 13.3 The Fabry-Perot interferometer.

---

*In the visible region of the spectrum, silver is the best metal to coat with (the reflectivity is about 0.97 in the red region and decreases to about 0.90 in the blue region). But beyond the blue region, the reflectivity falls rapidly. Aluminum is usually employed below 4000 Å.

each plate are made to have a slight angle (~ 1 to 10 minutes of arc) between them in order to avoid the unwanted fringes formed due to multiple reflections in the plate itself.

One of the two plates is kept fixed while the other can be moved to vary the separation of the two plates. In this configuration the instrument is called a Fabry-Perot interferometer. Sometimes both the plates are kept at a fixed separation (ranging from 1 to 200 mm) with the help of spacers. The system with fixed spacing is known as Fabry-Perot etalon. The Fabry-Perot interferometer (or etalon) is used to determine wavelengths precisely, to compare two wavelengths, to calibrate the standard meter in terms of wavelengths, etc.

Plate 9b shows the fringes obtained using a Fabry-Perot interferometer. Also shown in the figure for comparison, are the fringes obtained by using Michelson interferometer [see Plate 9a]. It can readily be seen that the Fabry-Perot interferometer, which employs the principle of multiple beam interference, produces much sharper fringes and hence can be used to study hyperfine structure of spectral lines.

13.3.1 Resolving Power of a Fabry-Perot Interferometer

The important consideration in a spectroscope is the ability of the instrument to resolve spectral lines close to each other. This is determined by the chromatic resolving power of the spectroscope which is defined as the ratio $\lambda/\Delta\lambda$, where the spectroscope just resolves the lines λ and $\lambda+\Delta\lambda$. There exist different criteria to specify whether two spectral lines can be said to be just resolved. The most widely used criterion is the Rayleigh criterion which will be discussed in Sec. 15.4. This criterion stated in the form as given in Sec. 15.4 cannot be applied here since the intensity never drops down to zero; secondary maxima and minima do not exist in the present case. But to be able to compare with other spectroscopes (like those employing diffraction gratings), we will again use the same criterion but stated in a slightly different form, namely that two wavelength components will be said to be just resolved when the intensity dip at the centre of the total pattern is 81.1% of the maximum (see Fig. 15.19). Using this criterion we will now calculate the chromatic resolving power of a Fabry-Perot interferometer.

In a Fabry-Perot interferometer it is the fringe pattern formed by transmitted light that is observed and as such the intensity distribution would be given by [see Eq. (13.2-9)]:

$$I = \frac{I_0}{1 + F\sin^2\delta/2} \qquad (13.3\text{-}1)$$

where I_0 is the incident intensity and

$$\delta = \frac{4\pi}{\lambda} h \cos\theta_r; \qquad (13.3\text{-}2)$$

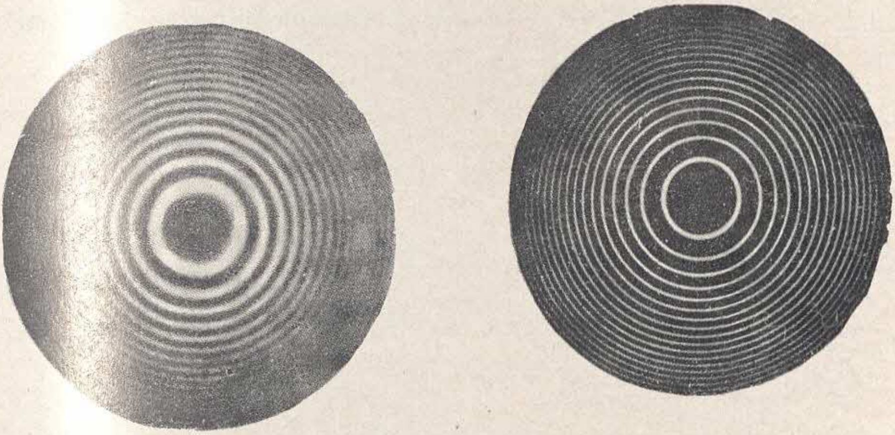


Plate 9 Fringes observed using (a) Michelson interferometer, (b) Fabry-Perot interferometer (From F. A. Jenkins and H. E. White, *Fundamentals of Optics*, McGraw-Hill Book Co. New York, 1976; used with permission.

here we have used the fact that $n \simeq 1$ since the medium between the two reflecting surfaces is air.* A change $\Delta\lambda$ in the wavelength would correspond to a change in δ given by

$$\Delta\delta = \frac{4\pi h \cos \theta_r}{\lambda^2} \Delta\lambda \tag{13.3-3}$$

where we have dropped the negative sign since it is not of any consequence in the following analysis. The total intensity distribution, resulting from the overlapping of the two intensity distributions peaked at λ and $\lambda+\Delta\lambda$, will be (see Fig. 13.4)

$$I_{\text{tot}} = \frac{I_0}{1+F\sin^2 \delta/2} + \frac{I_0}{1+F\sin^2\left(\frac{\delta-\Delta\delta}{2}\right)} \tag{13.3-4}$$

We will assume that the mxaima of I_{tot} to coincide with those of original maxima. Then, if the order of the fringe for λ is m, then $\delta = 2m\pi$. Now, the intensity at the mid point will correspond to $\delta = 2m\pi + \frac{\Delta\delta}{2}$. Thus,

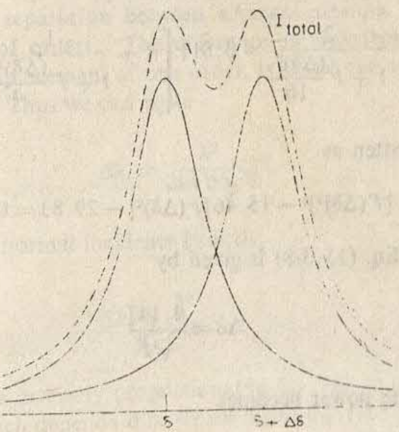


Fig. 13.4 The total intensity distribution at the Rayleigh resolution limit.

$$I_{\text{mid}} = \frac{I_0}{1+F\sin^2\left(m\pi+\frac{\Delta\delta}{4}\right)} + \frac{I_0}{1+F\sin^2\left(m\pi-\frac{\Delta\delta}{4}\right)} \simeq \frac{2I_0}{1+F\sin^2\frac{\Delta\delta}{4}} \tag{13.3-5}$$

---

*It must be noted here that the values of the wavelengths obtained using a Fabry-Perot interferometer will be air wavelengths unless the interferometer is kept in an evacuated chamber. Otherwise one must, in general, correct for the small dispersion of air.

and the intensity at the maximum will be

$$I = I_0 + \frac{I_0}{1 + F\sin^2 \frac{\Delta\delta}{2}} \tag{13.3-6}$$

Now according to the Rayleigh criterion the wavelengths λ and $\lambda + \Delta\lambda$ are just resolved if

$$I_{\text{mid}} = 0.811\ I_{\max}$$

or

$$\frac{2I_0}{1 + F\sin^2 \frac{\Delta\delta}{4}} = 0.811 \left[\frac{I_0}{1 + F\sin^2 \frac{\Delta\delta}{2}} + I_0 \right] \tag{13.3-7}$$

If the fringes are sharp enough, $\Delta\delta$ will be very small and we can replace $\sin \frac{\Delta\delta}{2}$ by $\frac{\Delta\delta}{2}$ (the error introduced will be less than 1%). Thus, we would obtain

$$\frac{2}{1 + F\frac{(\Delta\delta)^2}{16}} = 0.811 \left[1 + \frac{1}{1 + F\frac{(\Delta\delta)^2}{4}} \right]$$

which can be written as

$$[F(\Delta\delta)^2]^2 - 15.46[F(\Delta\delta)^2] - 29.83 = 0 \tag{13.3-8}$$

The solution of Eq. (13.3-8) is given by

$$\Delta\delta = \frac{4.147}{\sqrt{F}} \tag{13.3-9}$$

Thus the resolving power becomes

$$\begin{aligned} \frac{\lambda}{\Delta\lambda} &= \frac{4\pi h \cos\theta_r}{\lambda\Delta\delta} = \frac{4\pi h \cos\theta_r \sqrt{F}}{4.147\lambda} \\ &= \frac{2\pi\sqrt{F}}{4.147} m = 1.515 m\sqrt{F} \end{aligned} \tag{13.3-10}$$

To have an idea of the numerical value of resolving power, let us consider a Fabry-Perot etalon with $h = 1$ cm and $R = 0.8$ ($F = 80$). The resolving power for normal incidence in the wavelength region around $\lambda = 5000$ Å would be

$$\frac{\lambda}{\Delta\lambda} = \frac{4\pi\sqrt{80}}{5\times10^{-5}\times4.147} \simeq 5.42\times10^5 \tag{13.3-11}$$

that is, two wavelengths separated by 0.0092 Å can be resolved at

$\lambda = 5000$ Å. With $R = 0.9\,(F = 360)$, $\frac{\lambda}{\Delta\lambda} \simeq 11.5 \times 10^6$ and $\Delta\lambda \simeq 0.00446$ Å

Thus a Fabry-Perot instrument can resolve wavelengths differing by about 10^{-3} Å. This is in contrast to that of a grating (say having 25000 grooves) which resolves up to about 0.1 Å at $\lambda = 5000$ Å and that of a prism (made of dense flint glass with 5 cm base) which resolves only up to about 1 Å at 5000 Å.

It must be noted that in the above analysis, we have considered two monochromatic lines at λ and $\lambda + \Delta\lambda$. In general, the lines at the two wavelengths λ and $\lambda + \Delta\lambda$ themselves will be wide and this restricts the use of such high resolving powers.

When the Fabry-Perot interferometer is used to analyse spectra with very closely spaced lines, then the distance between the adjacent maxima would be greater than the displacement between the system of rings of the spectral lines. But when the spectrum has widely separated wavelength components, then it might happen that the displacement between the rings is greater than the separation between adjacent maxima. This results in the "overlapping" of orders. The difference in wavelength ($\Delta\lambda_s$) which corresponds to a displacement of one order, is called the spectral range of the interferometer. Thus we can write

$$\Delta\lambda_s = \frac{\lambda^2}{2nh\cos\theta} \tag{13.3-12}$$

This becomes, near normal incidence ($\theta \simeq 0$)

$$\Delta\lambda_s = \frac{\lambda^2}{2nh} \tag{13.3-13}$$

which is found to be inversely proportional to h. This is in contrast to the resolving power which depends directly on h [see Eq. (13.3-10)].

When the spectrum is complex consisting of a number of widely separated wavelength components each with a hyperfine structure, then one can separate the different wavelength components by employing the Fabry-Perot interferometer along with a spectrograph as shown in Fig. 13.5(a). The light emerging from the source S is rendered parallel by the lens L_1. The interference pattern formed by the Fabry-Perot interferometer (marked F.P. in the figure) is made to fall on the slit of the spectrograph. The spectrograph separates the spectral components and one obtains in the plane P images of the slit, each crossed by fringes as shown in Fig. 13.5(b).

13.4 The Lummer-Gehrcke Plate

We saw in Sec. 13.2 that the sharpness of fringes (and hence the resolving power) of a multiple beam interferometer increases as the reflectivity R of the plate increases. But one cannot use very thick coating of metals to increase the reflectivity as the intensity of the beam would be reduced considerably due to absorption in metallic coatings. This difficulty can be overcome by the use of the phenomenon of total internal reflection (instead of metallic reflection); this is used in the Lummer-Gehrcke plate which will be discussed in this section.

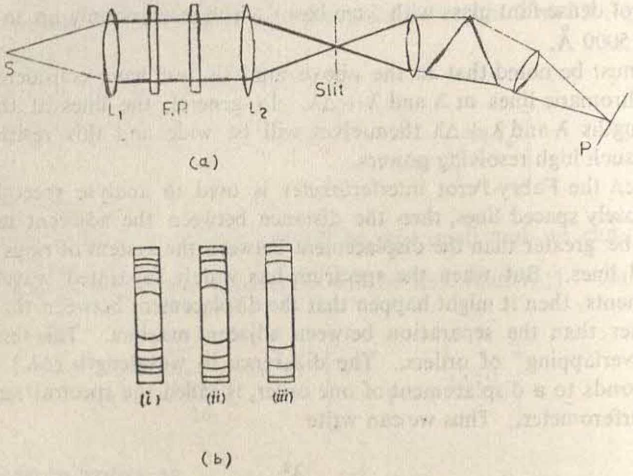


Fig. 13.5 (a) A Fabry-Perot interferometer used in conjunction with a spectrograph. (b) The interlaced fringes formed in the plane of the slit are separated by the prism. For example, (i), (ii) and (iii) may correspond to the lines in the red, yellow and green regions respectively as observed on plane *P*.

A Lummer-Gehrcke plate is a plane parallel plate made of glass (or quartz), on one end of which a small right angle prism of the same material is fixed (see Fig. 13.6). The angle of the prism is chosen in such a way

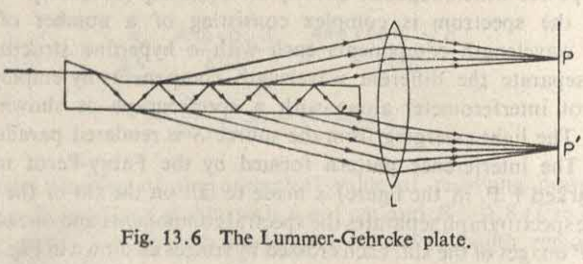


Fig. 13.6 The Lummer-Gehrcke plate.

that the rays incident normally on the surface of the prism hit the two surfaces of the plate at an angle slightly less than the critical angle.* Since the two surfaces are parallel, all successive reflections will occur at the same (near critical) angle. Most of the light will be reflected with a little fraction being transmitted at each reflection. Thus, there will emerge from the upper and lower surfaces of the plate a series of waves which would finally interfere to produce interference fringes in the plane P (see Fig. 13.6). Notice that the prism suppresses the externally reflected beam. In the plane P, one obtains fringe patterns on either side of the plate. The fringes are approximately straight lines parallel to the plate surfaces.

---

*Beyond the critical angle, the reflection is total (see Sec. 20.2) while slightly below the critical angle, the reflectivity is high.

We will not go into the details of the theory of the Lummer-Gehrcke plate but two points may be noted:

(i) Unlike in the case of Fabry-Perot interferometer, the space between the reflecting surfaces is a dispersive medium, and

(ii) The number of reflections is also not very large as in the case of the Fabry-Perot interferometer; the number of reflections depends on the length of the plate and the angle θ_r (see Fig. 13.6). Thus, the resolving power of the instrument depends on the length of the plate.

Earlier, Lummer-Gehrcke plates were used in high resolution spectroscopy. However, it has been replaced by the more flexible Fabry-Perot interferometer.

13.5 Interference Filters

When a Fabry-Perot interferometer is illuminated by a monochromatic (uncollimated) beam, we get a spectrum consisting of different intensity maxima which satisfy the following relation:

$$2nd\cos\theta_r = m\lambda \tag{13.5-1}$$

Now if a Fabry-Perot interferometer is illuminated with a collimated white light incident normally ($\theta_r \simeq 0$), maxima of different orders are formed in the transmitted light corresponding to wave lengths given by

$$\lambda = \frac{2nd}{m} \tag{13.5-2}$$

If d is large, a large number of maxima will be observed in the visible region; for example, about 23,000 maxima are observed if $d = 1$ cm. But, if we go on reducing d, we reach a situation in which only one or two maxima are obtained in the visible region. For example, if $n = 1.5$ and $d = 6 \times 10^{-3}$ cm, there are only two maxima in the visible region, corresponding to $\lambda = 6000$ Å ($m = 3$) and $\lambda = 4500$ Å ($m = 4$). They are widely separated and one of them can be masked so as to transmit only one wavelength. In this way, it is possible to filter a particular wavelength out of a white light beam. Such a structure is known as an interference filter.* Interference filters using this principle can be obtained by modern vacuum deposition techniques. A thin metallic film (usually, of aluminum or silver) is deposited on a substrate (generally, a glass plate) by vacuum deposition techniques. Then a thin layer of a dielectric material such as cryolite ($3\,NaF.AlF_3$) is deposited over this. This structure is again covered by another metallic film (see Fig. 13.7). To protect this film structure from any damage, another glass plate is placed over it. Thus a Fabry-Perot structure is formed between the two glass plates. By varying the thickness of the dielectric film, one can filter out any particular wavelength. However, the filtered light will have a finite width, that is, it will have a narrow spectrum sharply peaked about one wavelength. The sharpness of the transmitted spectrum is determined by there solving power of the formed Fabry-Perot structure, and hence by the reflectivity of the surfaces. The larger the reflectivity, the narrower is the transmitted spectrum. But it is not possible to increase the thickness of metallic films indefinitely as absorption will reduce the intensity of the transmitted light. To overcome this difficulty, metallic films are replaced by all dielectric structures.

---

*The Fabry-Perot structure also behaves as a resonator and supports the oscillation of what are known as modes (see Sec. 22.4).

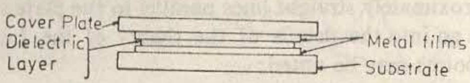


Fig. 13.7 The interference filter.

In an all-dielectric structure, layers of dielectric materials of appropriate refractive indices are deposited. It was shown in Chapter 12 how dielectric films can be used to enhance the reflectivity of a surface. If, on a glass-plate, a $\lambda/4$ thick film of a dielectric material whose refractive index is more than that of glass, is deposited, the reflectivity of the glass plate increases. Larger the difference between the refractive indices, greater will be the reflectivity. The materials generally used in interference filters are titanium oxide ($n=2.8$) or zinc sulphide ($n=2.3$). To obtain interference filters, a $\lambda/4$ thick film of titanium oxide is deposited on a glass substrate. Then a thin layer of dielectric material with lower refractive index (such as cryolite or magnesium fluoride) is deposited. Onthis is again deposited a $\lambda/4$ thick layer of a material of higher refractive index. To increase the reflectivity, multilayer structures of alternate higher and lower refractive index materials are used. In this way, it is possible to achieve a reflectivity of more than 90% for any particular wavelength. Such filters are capable of transmitting over a bandwidth as small as 11 Å or even less with peak at any wavelength within the visible region.

PROBLEMS

1. Derive Eq. (13.2-9).
2. Show that the half-width of the fringes formed with a Fabry-Perot interferometer

is given by Eq. (13.2–17). (Half-width is the full width of the fringe at half the maximum value of the intensity).

3. Calculate the resolving power of a Fabry-Perot interferometer made of reflecting surfaces of reflectivity 0.85 and separated by a distance 1 mm at $\lambda = 4880$ Å.
4. Calculate the minimum spacing between the plates of a Fabry-Perot interferometer which would resolve two lines with $\Delta\lambda = 0.1$ Å at $\lambda = 6000$ Å. Assume the reflectivity to be 0.8.
5. Show that in the system of fringes formed in transmitted light by multiple reflection (i.e., in the system of fringes formed in a Fabry-Perot interferometer), the ratio of the intensity of the maxima to that midway between two maxima is $(1 + R^2)^2/(1 - R^2)^2$. This is just the contrast.

REFERENCES AND SUGGESTED READING

Baker, A., (1970). *Modern Physics and Antiphysics*, Addison Wesley Publishing Co., Reading, Mass. (USA).

Baumeister, P. and Pincus, G., (1970), "Optical interference coatings," *Scient. Amer.* 223, December, 59.

Born, M. and Wolf, E., (1975), *Principles of Optics*, Pergamon Press, Oxford (England).

Cagnet, M., Francon, M., and Mallick, S., (1971), *Atlas of Optical Phenomena*, Springer-Verlag, Berlin (Germany).

Cagnet, M., Francon, M. and Thierr, J.C., (1962), *Atlas of Optical Phenomena*, Springer-Verlag, Berlin (Germany).

Cave, E.F. and Holroyd, L.V., (1955), "Inexpensive Michelson interferometer," *Amer. J. Phys.*, **23**, 61.

Cleveland, E.L., (1959), "Double slit interference," *Amer. J. Phys.*, **27**, 521.

Collett, E. (1971), "Mathematical formulation of the interference laws of Fresnel and Arago," *Amer. J. Phys.*, **39**, 1483.

Cook, A.H., (1971), *Interference of Electromagnetic Waves*, Clarendon Press, Oxford (England).

Ditchburn, R.W., (1963), *Light*, John Wiley and Sons, New York (USA).

Francon. M., (1963), *Modern Applications of Physical Optics*, Interscience, New York (USA).

Francon, M., (1966), *Optical Interferometry*, Academic Press, New York (USA).

Feynman, R.P., Leighton, R.B. and Sands, M., (1965), *The Feynman Lectures on Physics*, Vol. I, Addison Wesley Publishing Co., Reading, Mass, (USA).

Jenkins, F.A. and White, H.E., (1976), *Fundamentals of Optics*, McGraw-Hill Book Co., New York (USA).

Oppenheim, V. and Jaffe, J.H., (1956), "Interference in an optical wedge," *Amer. J. Phys.*, **24**, 610.

Sladkova, J., (1968), *Interference of Light*, Iliffe Books Ltd., London (England).

Steel, W.H., (1967), *Interferometry*, Cambridge University Press, London (England).

Tolansky, S., (1948), *Multiple Beam Interferometry of Surfaces and Films*, Oxford University Press, London (England).

Tolansky, S., (1955), *An Introduction to Interferometry*, Longmans Green and Co., London (England).

14 Coherence

A privacy of glorious light is thine,
Whence thou dost pour upon the world a flood
Of harmony, with instinct more divine.

WORDSWORTH

14.1 Temporal Coherence

The displacement corresponding to a one-dimensional plane wave, propagating in the $+x$-direction, can be written in the form

$$\Psi(x, t) = A\, e^{i(kx-\omega t)} \tag{14.1-1}$$

where A represents the amplitude of the wave which may in general be complex. If we write

$$A = |A|\, e^{i\varphi}, \tag{14.1-2}$$

then Eq. (14.1-1) becomes

$$\Psi(x, t) = |A|\, e^{i(kx-\omega t+\varphi)} \tag{14.1-3}$$

The actual displacement, which we will denote by E, is the real part of Ψ and is, therefore, given by

$$E = |A| \cos(kx-\omega t+\varphi) \tag{14.1-4}$$

However, the displacement represented by Eq. (14.1-4) is a practical impossibility because at an arbitrary value of x, Eq. (14.1-4) predicts that the displacement is sinusoidal for $-\infty < t < +\infty$ [see Fig. 14.1(a)]. For example

$$E(x=0, t) = |A| \cos(\omega t - \varphi) \tag{14.1-5}$$

In practice, for a typical light source (like a sodium lamp), the field is not perfectly sinusoidal for all values of time and it is indeed of the form as shown in Fig. 14.1(b). This is because of the fact that when an excited atom goes to the ground state and emits light, the duration of the pulse is of the order of 10^{-10} sec. Thus, the field remains sinusoidal for times of the order of 10^{-10} sec.* This is known as the *coherence time* of the optical beam and is usually denoted by τ_c. It may be noted that for an optical beam ($\nu \sim 5 \times 10^{14}$ sec^{-1}) and even for $t \simeq 10^{-10}$ sec one has about 5×10^4 oscillations.

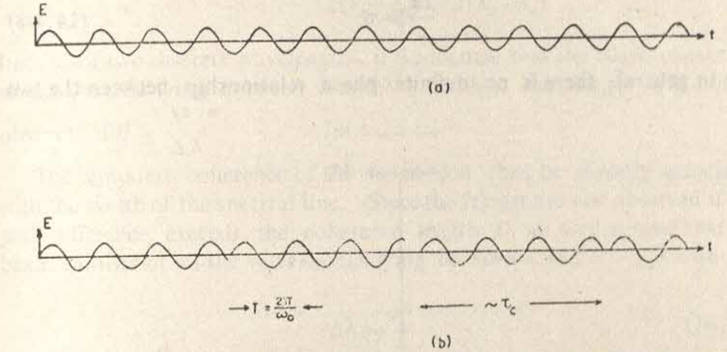


Fig. 14.1 (a) For a perfectly monochromatic beam, the displacement remains sinusoidal for $-\infty < t < \infty$.

(b) For an actual source definite phase relationship exist for times of the order of τ_c, which is known as the temporal coherence of the beam. For $\nu \sim 5 \times 10^{14}$ sec^{-1} and $\tau_c \sim 10^{-10}$ sec one has about 50000 oscillations.

The coherence length L, of the light is defined by the following equation

$$L = c\tau_c \tag{14.1-6}$$

where c represents the speed of light. In order to understand the concept of coherence time (or of coherence length) let us consider the Michelson interferometer experiment (see Sec. 12.10). A light beam falls on a beam splitter G (which is usually a partially silvered plate) and the waves reflected from the mirrors M_1 and M_2 interfere (see Fig. 14.2). Let M_2' represent the image of the mirror M_2 (formed by the plate G) as seen by the eye. If the distance M_1M_2' is denoted by d, then, the beam which gets reflected by the mirror M_2 travels an additional path equal to $2d$. Thus, the beam reflected from the mirror M_1 interferes with the beam reflected from the mirror M_2 which had originated $2d/c$ seconds earlier. If the distance d is such that

$$\frac{2d}{c} \ll \tau_c \tag{14.1-7}$$

---

*However, for a laser beam the field remains sinusoidal for extremely large times (see Sec. 14.2).

then definite phase relationship exist between the two beams and well defined interference fringes are observed. On the other hand, if

$$\frac{2d}{c} \gg \tau_c \tag{14.1-8}$$

then, in general, there is no definite phase relationship between the two

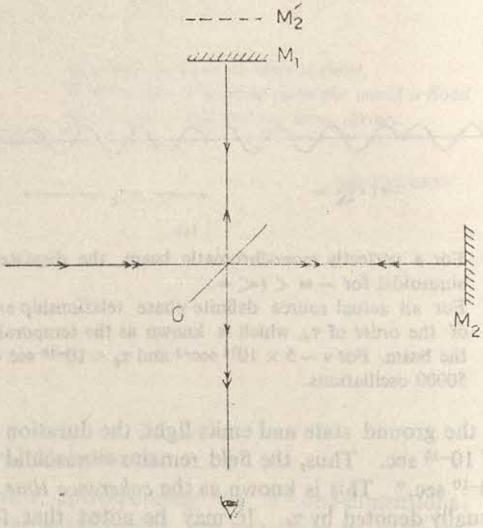


Fig. 14.2 The Michelson interferometer arrangement. G represents the beam splitter. M_2' represents the image of M_2 as formed by G.

beams and no interference pattern is observed. It may be mentioned that there is no definite distance at which the interference pattern disappears: as the distance increases, the contrast of the fringes becomes gradually poorer and eventually the fringe system disappears. For the neon 6328 Å line, the disappearance occurs when the path difference is about a few centimeters. Thus, $\tau_c \sim 10^{-10}$ sec. On the other hand, for the red cadmium line ($\lambda = 6438$ Å), the coherence length is of the order of 30 cm giving $\tau_c \sim 10^{-9}$ sec. For commercially available lasers, the coherence length can exceed a few kilometers.

14.2 The Line Width

In Sec. 12.10 we had shown that if we have two closely spaced wavelengths λ_1 and λ_2 (like the D_1 and D_2 lines of sodium), then in the Michelson interferometer experiment, the interference pattern will disappear if

$$\frac{2d}{\lambda_2} - \frac{2d}{\lambda_1} = \frac{1}{2} \tag{14.2-1}$$

where $2d$ represents the path difference between the two beams. Thus

$$2d = \frac{\lambda_1\lambda_2}{2(\lambda_1-\lambda_2)} \simeq \frac{\lambda^2}{2(\lambda_1-\lambda_2)} \tag{14.2-2}$$

Instead of two discrete wavelengths, if we assume that the beam consists of all wavelengths lying between λ and $\lambda + \Delta\lambda$ then the fringes will not be observed if $d \leq \frac{\lambda^2}{\Delta\lambda}$.*

The temporal coherence of the beam can thus be directly associated with the width of the spectral line. Since the fringes are not observed if the path difference exceeds the coherence length L, we may assume that the beam consists of all the wavelengths lying between λ and $\lambda + \triangle\lambda$ with

$$\Delta\lambda \sim \frac{\lambda^2}{L} \tag{14.2-3}$$

For the red cadmium line, $\lambda = 6.438 \times 10^{-5}$ cm and $L \simeq 30$ cm giving

$$\Delta\lambda \sim \frac{\lambda^2}{L} = \frac{(6.438\times10^{-5})^2}{30}$$

$$\sim 0.01\ \text{Å}$$

For the sodium line, $L \simeq 3$ cm and

$$\Delta\lambda \sim 0.1\ \text{Å}$$

Further, since $\nu = c/\lambda$, the frequency spread $\Delta\nu$ of a line would be

$$\Delta\nu \sim \frac{c}{\lambda^2}\Delta\lambda$$

$$\sim \frac{c}{L} \tag{14.2-4}$$

where we have disregraded the sign. Since $\tau_c \sim \frac{L}{c}$, we obtain

---

*This follows from the fact that if

$$2d = \frac{\lambda^2}{2(\triangle\lambda/2)} = \frac{\lambda^2}{\Delta\lambda}$$

then the interference pattern produced by the wavelengths λ and $\lambda + \frac{\Delta\lambda}{2}$ will disappear. For each wavelength (lying between λ and $+ \lambda\frac{\triangle\lambda}{2}$) there will be a corresponding wavelength $\left(\text{lying between } \lambda + \frac{\triangle\lambda}{2} \text{ and } \lambda + \triangle\lambda\right)$ such that the minima of one falls on the maxima of the other resulting in disappearance of the fringes. Clearly, for $d \gtrsim \frac{\lambda^2}{\triangle\lambda}$, the contrast will be extremely poor.

$$\Delta\nu \sim \frac{1}{\tau_c} \tag{14.2-5}$$

Thus, the frequency spread of a line is of the order of the inverse of the coherence time. This can also be seen by Fourier analysis. In order to show this, we consider a displacement which is finite only for a time τ_c. For example, we may have

$$\left.\begin{aligned} \Psi(x=0, t) &= a\, e^{i\omega_0 t} \quad |t| < \tau_c/2 \\ &= 0 \qquad\quad\; |t| > \tau_c/2 \end{aligned}\right\} \tag{14.2-6}$$

If we write

$$\Psi(x=0, t) = \int_{-\infty}^{\infty} A(\omega)\, e^{i\omega t}\, d\omega \tag{14.2-7}$$

then by using the Fourier integral theorem (see Sec. 7.4) we obtain

$$A(\omega) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \Psi(x=0, t)\, e^{-i\omega t}\, dt \tag{14.2-8}$$

In Eq. (14.2-7), $A(\omega)$ represents the amplitude[*] of the plane wave corresponding to the frequency ω. Substituting for $\Psi(x=0, t)$ from Eq. (14.2-6) in Eq. (14.2-8), we obtain

$$\begin{aligned} A(\omega) &= \frac{1}{2\pi}\int_{-\tau_c/2}^{+\tau_c/2} a\, e^{i(\omega_0-\omega)t}\, dt \\ &= \frac{a}{\pi}\left[\frac{\sin\left[\frac{1}{2}(\omega-\omega_0)\tau\right]}{(\omega-\omega_0)}\right] \end{aligned}$$

---

[*] Notice that the integral appearing on the RHS of Eq. (14.2-7) is over negative values of ω also. However, the displacement is the real part of Ψ which is given by

$$\begin{aligned} E = Re\,[\psi(x=0, t)] &= Re\left[\left[\int_{-\infty}^{\infty} |A(\omega)|\, e^{i(\omega t+\varphi)}\, d\omega\right]\right. \\ &= \int_{-\infty}^{\infty} |A(\omega)| \cos(\omega t+\varphi)\, d\omega = \int_{0}^{\infty} |A(\omega)| \cos(\omega t+\varphi(\omega))\, d\omega \\ &\qquad + \int_{0}^{\infty} |A(-\omega)| \cos(\omega t-\varphi(-\omega))\, d\omega \end{aligned}$$

where we have used the relation $A(\omega) = |A(\omega)| \exp[i\varphi(\omega)]$. The above equation can always be written in the form

$$\int_{0}^{\infty} C(\omega) \cos(\omega t+\theta)\, d\omega$$

Thus the amplitudes associated with the negative frequencies contribute essentially to the corresponding positive frequencies.

or

$$A(\omega) = \frac{a}{\pi\omega_0}\,\frac{\sin[\alpha(\Omega - 1)]}{(\Omega - 1)} \tag{14.2-9}$$

where $\Omega = \frac{\omega}{\omega_0}$ and $\alpha = \frac{\omega_0 \tau_c}{2}$. In Fig. 14.3 we have plotted the function

$$\frac{\sin[\alpha(\Omega - 1)]}{(\Omega - 1)}$$

as a function of Ω for $\alpha = 10$; one can see that the function is sharply peaked around $\Omega = 1$ and the width*

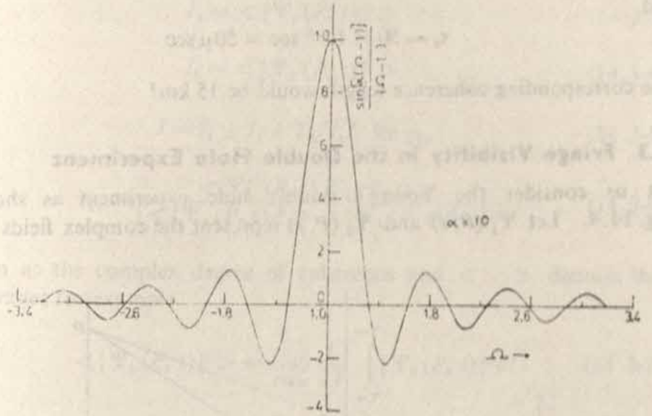


Fig. 14.3 The variation of the function $\frac{\sin(\Omega - 1)\alpha}{(\Omega - 1)}$ as a function of Ω for $\alpha = 10$.

$$\Delta\Omega\left(= \frac{\Delta\omega}{\omega_0}\right) \sim \frac{\pi}{\alpha} \tag{14.2-10}$$

Thus the function $A(\omega)$ is sharply peaked around $\omega = \omega_0$ and the width of $A(\omega)$ is given by

$$\Delta\,\omega \sim \frac{\pi\omega_0}{\alpha} = \frac{2\pi}{\tau_c}$$

or

$$\Delta\,\nu \sim \frac{1}{\tau_c} \tag{14.2-11}$$

consistent with Eq. (14.2-5). It may be noted that for a 0.1 nanosec optical pulse, (i.e., $\tau_c \sim 10^{-10}$ sec, $\omega_0 \simeq 3 \times 10^{15}$ sec^{-1}) $\alpha \simeq 1.5 \times 10^5$; thus, even for such a short pulse the value of α is very large and the corresponding frequency spread is very small.

---

*As can be seen from Fig. 14.3, the function becomes more sharply peaked as α increases.

As an example, for the 6328 Å line emitted from a neon source (at room temperature)*

$$\Delta\nu \sim 1.8 \times 10^{9} \text{ Hz}$$

Thus the coherence time

$$\tau_c \sim 0.5 \times 10^{-9} \text{ sec}$$

On the other hand, for a laser, line widths as low as 20 Hz have been obtained.** However, for a typical commercially available laser

$$\Delta\nu \sim 2 \times 10^{4}\ Hz$$

and

$$\tau_c \sim 50 \times 10^{-6} \text{ sec} = 50\,\mu\text{sec}$$

The corresponding coherence length would be 15 km!

14.3 Fringe Visibility in the Double Hole Experiment

Let us consider the Young's double hole experiment as shown in Fig. 14.4. Let $\Psi_1(P, t)$ and $\Psi_2(P, t)$ represent the complex fields at the point P due to the waves emanating from the holes S_1 and S_2 respectively. The resultant displacement would be given by

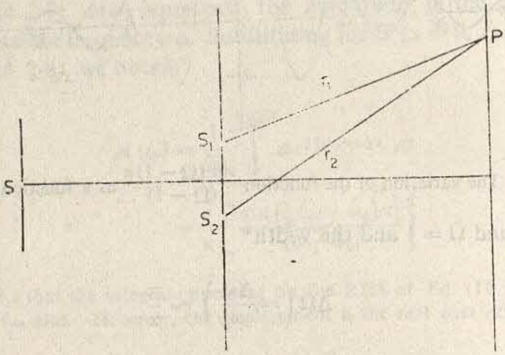


Fig. 14.4 Young's double hole experiment. The interference pattern observed around the point P at time t are due to the superposition of waves emanating from S_1 and S_2 at times $t-\frac{r_1}{c}$ and $t-\frac{r_2}{c}$ respectively.

$$\Psi = \Psi_1(P, t) + \Psi_2(P, t) \tag{14.3-1}$$

---

*This linewidth is primarily due to the Doppler broadening (see, for example, Meiners, 1970).

T.S. Jaseja, A. Javan and C.H. Townes, (1963), "Frequency Stability of He-Ne Masers and Measurement of Length", *Phys. Rev. Letters*, **10, 165.

Now, the intensity at the point P will be proportional to $|\Psi|^2$ which is given by the following equation

$$\begin{aligned}|\Psi|^2 &= \Psi_1^* \Psi_1 + \Psi_2^* \Psi_2 + \Psi_1^* \Psi_2 + \Psi_1 \Psi_2^* \\ &= |\Psi_1|^2 + |\Psi_2|^2 + 2\text{Re}\,(\Psi_1^* \Psi_2) \qquad (14.3\text{-}2)\end{aligned}$$

Since Ψ_1 and Ψ_2 vary with extreme rapidity, we can observe only the average value of $|\Psi^2|$. Thus, if we write

$$I_1 = <|\Psi_1(P, t)|^2> \qquad (14.3\text{-}3)$$

and

$$I_2 = <|\Psi_2(P, t)|^2> \qquad (14.3\text{-}4)$$

then

$$I = I_1 + I_2 + 2\sqrt{I_1 I_2}\ \text{Re}\ \gamma_{12} \qquad (14.3\text{-}5)$$

where

$$\gamma_{12} = \frac{<\Psi_1^*(P.\ t)\,\Psi_2(P, t)>}{[<|\Psi_1(P, t)|^2> <|\Psi_2(P, t)|^2>]^{1/2}} \qquad (14.3\text{-}6)$$

is known as the complex degree of coherence and $<\ldots>$ denotes the time average; for example

$$<|\Psi_1(P, t)|^2> = \lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{+T} |\Psi_1(P, t)|^2\, dt \qquad (14.3\text{-}7)$$

etc. The field $\Psi_1(P, t)$ is due to the waves emanating from the point S_1 at time $\left(t - \frac{r_1}{c}\right)$ where $r_1 = S_1P$. Thus, $\Psi_1(P, t)$ will be proportional to $\Psi\left(S_1, t - \frac{r_1}{c}\right)$ where $\Psi(S_1, t)$ denotes the field at S_1 at time t. Similarly $\Psi_2(P, t)$ will be proportional to $\Psi\left(S_2, t - \frac{r_2}{c}\right)$. Thus,

$$\gamma_{12} = \frac{<\Psi^*\left(S_1, t - \frac{r_1}{c}\right)\Psi^*\left(S_2, t - \frac{r_2}{c}\right)>}{\left[<\left|\Psi\left(S_1, t - \frac{r_1}{c}\right)\right|^2> <\left|\Psi\left(S_2, t - \frac{r_2}{c}\right)\right|^2>\right]^{1/2}} \qquad (14.3\text{-}8)$$

Since the overall intensity distribution in the fringe pattern does not change with time, we may write

$$\gamma_{12} = \frac{<\Psi^*(S_1, t + \tau)\,\Psi(S_2, t)>}{[<|\Psi(S_1, t)|^2> <|\Psi(S_2, t)|^2>]^{1/2}} \qquad (14.3\text{-}9)$$

where $\tau = \frac{r_2 - r_1}{c}$. In order to discuss the effect of temporal coherence, we may assume S, S_1 and S_2 to be of negligible spatial dimension

Further, if S_1 and S_2 are equidistant from the point S, then, we may assume

$$\Psi(S_1, t) = \Psi(S_2, t) = \Psi(t) \tag{14.3-10}$$

Thus, for such a case

$$\gamma_{12}(\tau) = \frac{<\Psi^*(t+\tau)\Psi(t)>}{<|\Psi(t)|^2>} \tag{14.3-11}$$

We write

$$\Psi(t) = A(t)\, e^{-i[\omega t+\varphi(t)]} \tag{14.3-12}$$

where $A(t)$ and $\varphi(t)$ are slowly varying real functions of time. For a perfectly monochromatic beam (i.e., infinite coherence time) $A(t)$ and $\varphi(t)$ are constants [see Fig. 14.1(a)] and

$$\Psi^*(t+\tau)\,\Psi(t) = A^2 e^{i\omega\tau} \tag{14.3-13}$$

Consequently

$$\gamma_{12}(\tau) = e^{i\omega\tau} \tag{14.3-14}$$

Thus, for such a case

$$I = I_1 + I_2 + 2\sqrt{I_1 I_2}\cos\omega\tau \tag{14.3-15}$$

and the visibility V, which is defined by

$$V = \frac{I_{\max} - I_{\min}}{I_{\max} + I_{\min}} \tag{14.3-16}$$

would be given by

$$V = \frac{2\sqrt{I_1 I_2}}{I_1 + I_2} \tag{14.3-17}$$

On the other hand, for an ordinary light source having $\tau_c \sim 10^{-10}$ sec, the functions $A(t)$ and $\varphi(t)$ can be assumed to be constants in times $\gtrsim 10^{-10}$ sec. Thus, if $\tau \gtrsim 10^{-10}$ sec, $\Psi(t+\tau)$ will have no phase relationship with $\Psi(t)$ and the time average $<\Psi^*(t+\tau)\Psi(t)>$ will vanish. Thus, if the path difference $S_2P \sim S_1P$ is such that

$$\frac{S_2P \sim S_1P}{c} \gtrsim \tau_c, \tag{14.3-18}$$

the fringe pattern will not be observed.

In general, we may write

$$\gamma_{12} = |\gamma_{12}|\, e^{i(\omega\tau+\beta)} \tag{14.3-19}$$

where $|\gamma_{12}|$ and β may be assumed to be constant around the observation point.* This gives us

---

*Notice from Eq. (14.3-14) that for a perfectly monochromatic beam, $|\gamma_{12}| = 1$ and $\alpha = \omega\tau = \frac{\omega}{c}(S_2P \sim S_1P)$. In general, it can be shown that $0 < |\gamma_{12}| < 1$; $|\gamma_{12}| = 0$ implies complete incoherence and $|\gamma_{12}| = 1$ implies complete coherence. In practice,

(*Contd.*)

$$I = I_1 + I_2 + 2\sqrt{I_1 I_2}\,|\gamma_{12}|\cos\alpha \tag{14.3-20}$$

where $\alpha = \omega\tau + \beta$. Thus

$$I_{\max} = I_1 + I_2 + 2\sqrt{I_1 I_2}\,|\gamma_{12}| \tag{14.3-21}$$

and

$$I_{\min} = I_1 + I_2 - 2\sqrt{I_1 I_2}\,|\gamma_{12}| \tag{14.3-22}$$

Hence, the visibility becomes

$$V = \frac{I_{\max} - I_{\min}}{I_{\max} + I_{\min}} = \frac{2\sqrt{I_1 I_2}}{I_1 + I_2}\,|\gamma_{12}| \tag{14.3-23}$$

Thus, the visibility is a direct measure of $|\gamma_{12}|$. In the present case, since S, S_1 and S_2 have been assumed to be point sources, $|\gamma_{12}|$ depends only on the temporal coherence of the beam. For $\tau << \tau_c$, $|\gamma_{12}|$ is very close to unity and for $\tau >> \tau_c$, $|\gamma_{12}|$ is close to zero.

14.4 Spatial Coherence

Till now, we have discussed the coherence of two fields arriving at a particular point P through two different optical paths. We will now consider the phase relationship between the fields at different points in space. Let us consider waves emanating from a point source. The points P and Q are equidistant from S (see Fig. 14.5). The point P' is along the line joining S and P. It is obvious that the phase relationship between

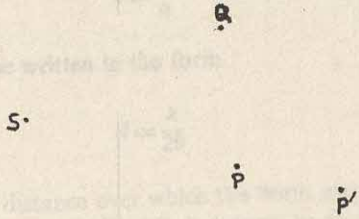


Fig. 14.5 For a point source S, definite phase relationship will always exist between the points P and Q.

if $|\gamma_{12}| > 0.88$, the light is said to be "almost coherent". Further, since

$$< \Psi^*(t+\tau)\,\Psi(t) > = e^{i\omega\tau} < A(t+\tau)\,A(t)\,e^{i[\varphi(t+\tau)-\varphi(t)]} >$$

and for a nearly monochromatic source $A(t)$ and $\varphi(t)$ are slowly varying functions of time, the quantity inside the angular brackets (on the RHS of the above equation) will not vary rapidly with τ. Thus, we may write $\gamma_{12} = |\gamma_{12}|\,e^{i\beta}\,e^{i\omega\tau}$ where both $|\gamma_{12}|$ and β are slowly varying functions of $\tau \left(= \dfrac{S_2P \sim S_1P}{c}\right)$.

the points P and P' will depend on the distance PP' and on the temporal coherence of the beam. If

$$\frac{PP'}{c} \gg \tau_c \tag{14.4-1}$$

no definite phase relationship will exist and conversely. On the other hand, the points P and Q will always have the same phase. However, if we have an extended source at S, then definite phase relationship may or may not exist between the points P and Q. In order to study the effect of the spatial extent of the source, let us consider the Young's double-hole experiment with two independent point source S and S' which are separated by a distance l [see Fig. 14.6(a)]. Each source will produce its own interference pattern. Let us assume that $SS_1 = SS_2$ and the point O is such that $S_1O = S_2O$. Clearly, the point source S will produce a

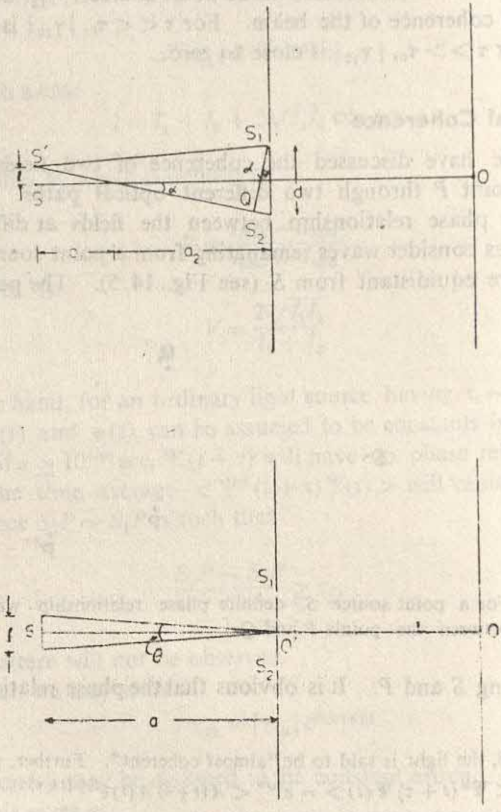


Fig. 14.6 (a) Young's double hole experiment with two independent point sources S and S'. (b) The same experiment with an extended source.

maximum around the point O. On the other hand, the intensity at the point O due to the point source S' will depend on the optical path $S'S_2 \sim S'S_1$. If

$$S'S_2 \sim S'S_1 = \tfrac{1}{2}\lambda \tag{14.4-2}$$

then the minima of the interference pattern produced by S will fall on the maxima of the interference pattern produced by S' and consequently no interference pattern will be observed. It can be easily seen that*

$$S'S_2 - S'S_1 \simeq \frac{ld}{a} \tag{14.4-3}$$

Thus no interference fringes will be observed if

$$\frac{ld}{a} = \tfrac{1}{2}\lambda \tag{14.4-4}$$

or

$$l = \frac{\lambda a}{2d} \tag{14.4-5}$$

Now, if we have an extended source [see Fig. 14.6(b)] whose linear dimension is $\sim \lambda a/2d$, then no interference pattern will be observed on the screen.

If θ represents the angle subtended by the source at the point O', then

$$\theta \simeq \frac{l}{a} \tag{14.4-6}$$

and Eq. (14.4-5) can be written in the form

$$d \simeq \frac{\lambda}{2\theta} \tag{14.4-7}$$

which gives the lateral distance over which the beam may be assumed to be spatially coherent. The quantity λ/θ is known as the lateral coherence width and is denoted by l_w.

*From Fig. 14.6a, we see that

$$S'S_2 - S'S_1 \approx S_2Q \approx \alpha d$$

But

$$\alpha \approx \frac{d/2}{a_2} = \frac{l}{a_1}$$

Thus

$$a = a_1 + a_2 \approx \left(l + \frac{d}{2}\right)\frac{1}{\alpha}$$

or

$$\alpha = \frac{l + d/2}{a}$$

Therefore

$$S'S_2 - S'S_1 \approx \alpha d \approx \left(l + \frac{d}{2}\right)\frac{d}{a} \approx \frac{ld}{a}$$

where we have neglected the term $d^2/2a$.

Example 14.1 For a circular source, the lateral coherence width is given by*

$$l_w = \frac{1.22\lambda}{\theta} \tag{14.4-8}$$

Consider a source S of 0.2 cm diameter [see Fig. 14.6(b)]. What should be the distance between S_1 and S_2 to obtain clear interference fringes. [Assume $\lambda \approx 5\times10^{-5}$ cm, $SS_1 \approx SS_2 \approx 1000$ cm].

Solution.

$$\theta \approx \frac{0.2\ \text{cm}}{1000\ \text{cm}} = 2\times10^{-4}\ \text{radians}$$

Thus

$$l_w \approx \frac{1.22\lambda}{\theta} \approx \frac{1.22\times5\times10^{-5}}{2\times10^{-4}} \approx 0.3\ \text{cm}$$

Consequently, the distance between S_1 and S_2 should be much less than 0.3 cm to obtain clear interference fringes.

14.5 Michelson's Stellar Interferometer

Using the concept of spatial coherence, one can determine the angular diameters of stars. The method is based on the principle that for a source of given spatial extent, as the distance between the slits S_1 and S_2 is increased, the contrast of the fringes (formed at the focal plane of the lens – see Fig. 14.7)** becomes poorer and eventually the fringes disappear. For a star whose angular diameter is $\sim 10^{-7}$ radians, the distance d for which the fringes will disappear would be

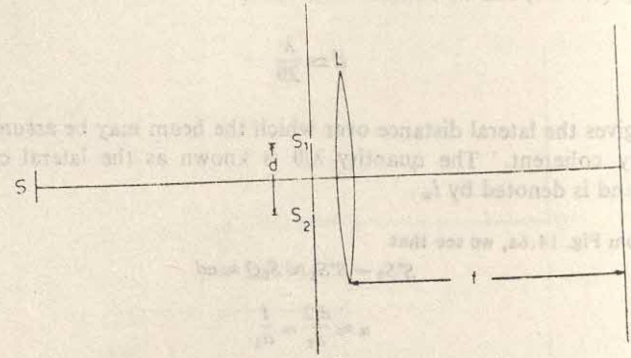


Fig. 14.7 S is a source of certain spatial extent; S_1 and S_2 are two slits separated by a distance d which can be varied. The fringes are observed on the focal plane of the lens L.

---

*This relation is derived in Born and Wolf (1975), Sec. 10.4.2.

**At the focal plane of the lens, one obtains the double slit diffraction pattern (see Sec. 15.5).

$$d \sim \frac{6\times10^{-5}}{\theta} = 600 \text{ cm}$$

Obviously, for such a large value of d, the fringe width will become extremely small. Further, one has to use a big lens, which is not only difficult to make, but only a small portion of it is used. Michelson used an ingenious technique to overcome these difficulties; he effectively got a large value of d by using two moveable mirrors M_1 and M_2 as shown in Fig. 14.8.

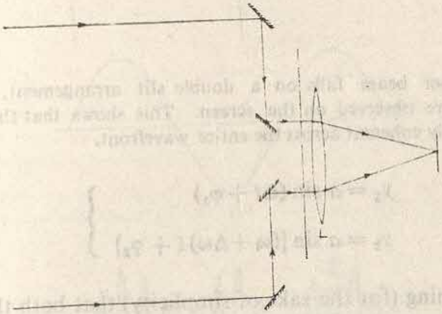


Fig. 14.8 Michelson's stellar interferometer.

The apparatus is known as Michelson's stellar inrteferometer. In a typical experiment the first disappearance occurred when the distance M_1M_2 was about 24 feet, which gave

$$\begin{aligned}\theta &\sim \frac{1.22\times5\times10^{-5}}{24\times12\times2.54} \text{ radians}\\ &\simeq 0.02''\end{aligned} \tag{14.5-1}$$

for the angular diameter of the star. This star is known as Arctures. From the known distance of the star one could estimate that the diameter of the star was about 27 times that of the sun.

It may be worthwhile to mention that a laser beam is spatially coherent across the entire cross-section. Thus, if a laser beam is allowed to fall directly on a double slit arrangement (see Fig. 14.9), then as long as the beam falls on both the slits, clear interference pattern is observed on the screen. This shows that the laser beam is spatially coherent across the entire wavefront.

14.6 Optical Beats

When two tuning forks, one having a frequency of 256 vib/sec and the other having a frequency of 260 vib/sec are made to vibrate at the same time, then one hears a frequency of about 258 vib/sec whose intensity varies from zero to maximum and back with a frequency of 4 cps. This phenomenon is known as beats. It can be easily understood by considering the superposition of two waves having frequencies ω and $\omega+\Delta\omega$:

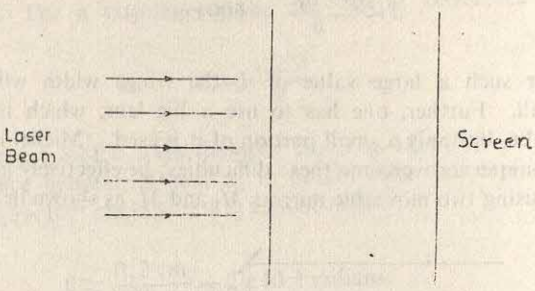


Fig. 14.9 If a laser beam falls on a double slit arrangement, interference fringes are observed on the screen. This shows that the laser beam is spatially coherent across the entire wavefront.

$$\left.\begin{aligned} y_1 &= a \sin(\omega t + \varphi_1) \\ y_2 &= a \sin[(\omega + \Delta\omega) t + \varphi_2] \end{aligned}\right\} \qquad (14.6\text{-}1)$$

where we are assuming (for the sake of simplicity) that both the waves have the same amplitude. The resultant displacement would be given by

$$\begin{aligned} y = y_1 + y_2 &= a[\sin(\omega t + \varphi_1) + \sin([\omega + \Delta\omega] t + \varphi_2)] \\ &= 2a \sin\left[\left(\omega + \frac{\Delta\omega}{2}\right) t + \frac{\varphi_1 + \varphi_2}{2}\right] \cdot \sin\left(\frac{\Delta\omega}{2} t + \frac{\varphi_1 - \varphi_2}{2}\right) \end{aligned} \qquad (14.6\text{-}2)$$

Figs. 14.10 (a) and 14.10 (b) show the time variations of the terms

$$\sin\left(\omega + \frac{\Delta\omega}{2}\right) t \quad \text{and} \quad \sin \frac{\Delta\omega}{2} t$$

respectively.* In Fig. 14.10 (c) we have plotted their product which represents the resultant displacement. Notice that although the envelope has a frequency of $\frac{\Delta\omega}{4\pi}\left(= \frac{\Delta\nu}{2}\right)$ [see Fig. 14.10 (b)] the intensity repeats itself after every $1/\Delta\nu$ seconds. This waxing and waning of sound is known as beats.

In a similar manner, one can consider the phenomenon of optical beats. For example, let us consider the superposition of two fields $\mathbf{E}_1$ and $\mathbf{E}_2$ having frequencies ω and $\omega + \Delta\omega$:

and

$$\left.\begin{aligned} \mathbf{E}_1 &= \mathbf{E}_{01} \sin(\omega t + \varphi_1) \\ \mathbf{E}_2 &= \mathbf{E}_{02} \sin([\omega + \Delta\omega] t + \varphi_2) \end{aligned}\right\} \qquad (14.6\text{-}3)$$

---

*We are assuming, without any loss of generality, $\varphi_1 = 0 = \varphi_2$.

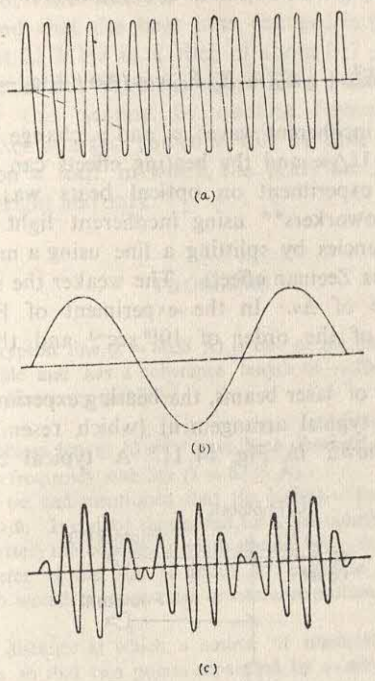


Fig. 14.10 (a) and (b) show typical time variations of sin $\left[\left\{\omega + \frac{\Delta\omega}{2}\right\} t\right]$ and sin $\frac{\Delta\omega}{2} t$ respectively; in (c) the time dependence of their product is shown.

If we assume that both the fields are linearly polarized in the same direction then in order to calculate the resultant field, we may simply algebraically add $\mathbf{E}_1$ and $\mathbf{E}_2$. Thus the resultant would be given by

$$\begin{aligned} E &= E_1 + E_2 \\ &= E_{01} \sin(\omega t + \varphi_1) + E_{02} \sin([\omega + \Delta\omega]\, t + \varphi_2) \end{aligned} \tag{14.6-4}$$

Now

$$\begin{aligned} E^2(t) = E_{01}^2 \sin^2(\omega t + \varphi_1) &+ E_{02}^2 \sin^2([\omega + \Delta\omega]\, t + \varphi_2) \\ &+ E_{01} E_{02} [-\cos(2\omega t + \Delta\omega\, t + \varphi_2 + \varphi_1) \\ &+ \cos(\Delta\omega\, t + \varphi_2 - \varphi_1)] \end{aligned} \tag{14.6-5}$$

For optical frequencies $\omega \approx 10^{15}$ sec^{-1} and the first three terms would vary with extreme rapidity. If we carry out an averaging over times which are long compared to $2\pi/\omega$ but short compared to $2\pi/\Delta\omega$, then we would obtain*

$$\langle E^2(t)\rangle = \tfrac{1}{2}E_{01}^2 + \tfrac{1}{2}E_{02}^2 + E_{01}E_{02}\cos(\Delta\omega t + \varphi_2 - \varphi_1) \qquad (14.6\text{-}6)$$

Clearly, for typical incoherent waves, φ_1 and φ_2 change in times which are small compared to $1/\Delta\omega$ and the beating effects can not be observed. However, the first experiment on optical beats was carried out by Forrester and his coworkers** using incoherent light. They used two closely spaced frequencies by splitting a line using a magnetic field (this splitting is known as Zeeman effect). The weaker the magnetic field, the smaller is the value of $\Delta\nu$. In the experiment of Forrester and his coworkers, $\Delta\nu$ was of the order of $10^{10}\ \text{sec}^{-1}$ and they were able to observe optical beats.

With the advent of laser beams, the beating experiments have become much easier and a typical arrangement (which resembles a Michelson interferometer) is shown in Fig. 14.11. A typical beat note of the

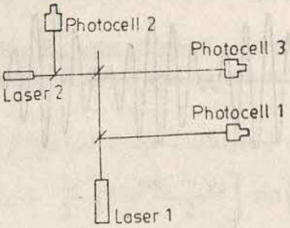


Fig. 14.11 The experimental arrangement of Lipsett and Mandel to observe optical beats using two laser beams.

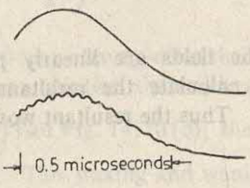


Fig. 14.12 Oscilloscope trace of the sum of the intensities of the laser beams (upper curve) and the intensity of the superposed laser beam (lower curve). [From Lipsett and Mandel (1963)].

---

*For example if $\Delta\nu \approx 20 \times 10^6$ Hz, then the time of averaging is small compared to 0.5×10^{-7} sec but large when compared to 5×10^{-14} sec. Thus if the detector resolution is about 10^{-9} sec then it will record only the average values of the first three terms of Eq. (14.6-5) and it will be able to record the time variation of the last term provided φ_1 and φ_2 do not change rapidly with time.

A.T. Forrester, R.A Gudmundsen and P.O. Johnson (1955), 'Photoelectric mixing of Incoherent Light', *Phys. Rev.* **99, 691.

experiment of Lipsett and Mandel* is shown in Fig. 14.12.

It was observed that the beat note changed in frequency from about 33 MHz to about 21 MHz in a time of about 0.7 μ sec. The coherence time is ~ 0.5 μ sec which is consistent with the duration of the spike.

We conclude this section by quoting Feynman that "with the availability of laser sources, someone will be able to demonstrate two sources shining on a wall, in which the beats are so slow that one can see the wall get bright and dark."

PROBLEMS

1. The orange Krypton line ($\lambda = 6058$ Å) is one of the most ideal monochromatic sources available and has a coherence length of ~ 20 cm. Calculate the line width and the frequency stability.
   (Ans. ~ 0.01 Å, $\sim 3 \times 10^{-6}$)
2. Laser line widths as low as 20 sec^{-1} have been obtained. Calculate the coherence length and the frequency stability ($\lambda = 6328$ Å).
3. In Sec. 14.4 we had mentioned that the lateral coherence width of a circular source is $1.22\lambda/\theta$. It can be shown that for good coherence (i.e., for a visibility of 0.88 or better) the coherence width should be $\leq 0.3\lambda/\theta$. Assuming that the angular diameter of the sun is about 30′, calculate the distance between two pinholes which would produce clear interference pattern.
   (Ans. ~ 0.02 mm).
4. Calculate the distance at which a source of diameter 1 mm should be kept from a screen so that two points separated by a distance of 0.5 mm may be said to be coherent. Assume $\lambda = 6 \times 10^{-5}$ cm.
5. In a Michelson interferometer experiment, it is found that for a source *S*, as one of the mirrors is moved away from equal path length position by a distance of about 5 cm, the fringes disappear. What is the coherence time of the radiation emerging from the source?
6. If we perform the Young's double hole experiment using white light, then only a few coloured fringes are visible. Assuming that the visible spectrum extends from 4000 to 7000 Å, explain this phenomenon qualitatively on the basis of coherence length.
7. Using the Stellar interferometer, Michelson observed for the star Betelgeuse, that the fringes disappear when the distance between the movable mirrors is 25 inches. Assuming $\lambda \simeq 6 \times 10^{-5}$ cm, calculate the angular diameter of the star.
8. Consider the Young's double hole experiment as shown in Fig. 11.5. The distance $SS_1 \simeq 1$m. Calculate the angular diameter of the hole *S* which would produce good interference pattern on the screen. Assume $\lambda = 6000$ Å.
9. Assume a Gaussian pulse of the form

$$\Psi(x=0,t) = A \exp\left(-\frac{t^2}{\sigma^2}\right) e^{-i\omega_0 t}$$

where $\sigma >>> \frac{1}{\omega_0}$. Plot the Fourier transform (as a function of ω) and interpret it physically. Show that the temporal coherence is $\sim \sigma$. What is the frequency spread?

---

*M.S. Lipsett and L. Mandel, (1963), "Coherence time Measurements of Light from Ruby Optical Masers', *Nature, Lond.*, **199**, 553.

10. In the above problem, assume $\lambda_0 = 6 \times 10^{-5}$ cm and $\sigma \sim 10^{-9}$ sec. Calculate the frequency components predominantly present in the pulse and compare it with the case corresponding to $\sigma \sim 10^{-6}$ sec.
11. In an experiment on optical beats, it was found that the beat frequency changed from about 10^8 Hz to 8×10^8 Hz in 1 microsecond. Explain this phenomenon and estimate the coherence time of the laser beam.

REFERENCES AND SUGGESTED READING

BORN, M. AND WOLF, E., (1975), *Principles of Optics*, Pergamon Press, Oxford (England).

FRANCON, M., (1966), *Diffraction: Coherence in Optics*, Pergamon Press, Oxford (England).

FRANCON, M., (1966), *Optical Interferometry*, Academic Press, New York (USA).

FORRESTER, A.T., (1956), On coherence properties of light waves, *Amer. J. Phys.*, **24**, 192.

FORRESTER, A.T., GUDMUNDSEN, R.A. AND JOHNSON, P.O., (1955), Photoelectric mixing of incoherent light, *Phys. Rev.*, **99** (6), 1891.

JASEJA, T.S., JAVAN, A. AND TOWNES, C.H., (1963), Frequency stability of He-Ne masers and measurement of length, *Phys. Rev. Lett.*, **10**, 165.

LIPSETT, M.S. AND MANDEL, L., (1963), Coherence time measurement of light from Ruby Optical masers, *Nature, Lond.*, **199**, 553.

MANDEL, L. AND WOLF, E., (1965), Coherence properties of optical fields, *Rev. Mod. Phys.*, **37**, 231.

MEINERS, H.F., (1970), *Physics Demonstration and Experiments*, Vol. 2, The Ronald Press Co., New York (USA).

WELTIN, H., (1962), Light beats, *Amer. J. Phys.*, **30**, 653.

PART IV

Diffraction

15 Fraunhofer Diffraction

At last the thin and shadowy form with noiseless,
trackless, footsteps came,

SHAKESPEARE

15.1 Introduction

Consider a plane wave incident on a long narrow slit of width b (see Fig. 15.1). According to geometrical optics one expects the region AB of the

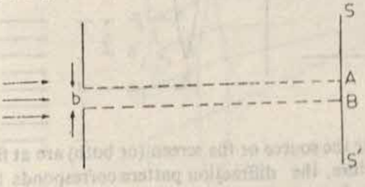


Fig. 15.1 If a plane wave is incident on an aperture then according to geometrical optics a sharp shadow will be cast in the region AB of the screen.

screen SS' to be illuminated and the remaining portion (known as the geometrical shadow) to be absolutely dark. However, if the observations are made carefully then one finds that if the width of the slit is not very large compared to the wavelength, then the light intensity in the region AB is not uniform and there is also some intensity inside the geometrical shadow. Further, if the width of the slit is made smaller, larger amount of energy reaches the geometrical shadow. This spreading out of a wave when it passes through a narrow opening is known as the phenomenon of diffraction and the intensity distribution on the screen is known as the diffraction pattern. We will discuss the phenomenon of diffraction in this chapter and will show that the spreading out decreases with decrease in wavelength. Indeed, since the light wavelengths are very small ($\sim 5\times10^{-5}$ cm), the effects due to diffraction are not readily observed.

The diffraction phenomena are usually divided into two categories: (i) Fresnel diffraction and (ii) Fraunhofer diffraction.

In the Fresnel class of diffraction the source of light and the screen are, in general, at a finite distance from the diffracting aperture [see Fig. 15.2(a)]. In the Fraunhofer class of diffraction, the source and the screen are at infinite distances from the aperture; this is easily achieved by placing the source on the focal plane of a convex lens and placing the screen on the focal plane of another convex lens [see Fig. 15.2 (b)]. The two lenses effectively move the source and the screen to infinity because the first lens makes the light beam parallel and the second lens makes the screen receive a parallel beam of light. It turns out that it is much easier to calculate the intensity distribution of a Fraunhofer diffraction pattern which we plan to do in this chapter.* Further, the Fraunhofer diffraction pattern is not

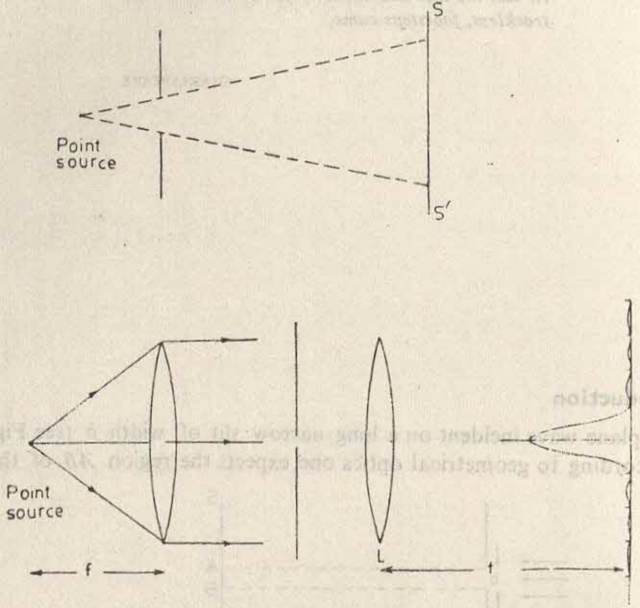


Fig. 15.2 (a) When either the source or the screen (or both) are at finite distances from the aperture, the diffraction pattern corresponds to the Fresnel class. (b) In the Fraunhofer class both the source and the screen are at infinity.

---

*In the next chapter we will study the Fresnel class of diffraction and will discuss the transition from the Fresnel region to the Fraunhofer region. We will also give a more general theory of Fraunhofer diffraction which will enable us to calculate the intensity distribution for various shapes of the aperture.

difficult to observe; all that one needs is an ordinary laboratory spectrometer; the collimator renders a parallel beam of light and the telescope receives parallel beams of light on its focal plane. The diffracting aperture is placed on the prism table.

15.2 Single Slit Diffraction Pattern

We will first study the Fraunhofer diffraction pattern produced by an infinitely long slit of width b. A plane wave is assumed to fall normally on the slit and we wish to calculate the intensity distribution on the focal plane of the lens L [see Fig. 15.3(a)]. We assume that the slit consists of a large number of equally spaced point sources and that each point on the slit is a source of Huygens' secondary wavelets which interfere with the wavelets emanating from other points. Let the point sources be at A_1, A_2, A_3, ...and let the distance between two consecutive points be Δ [see Fig

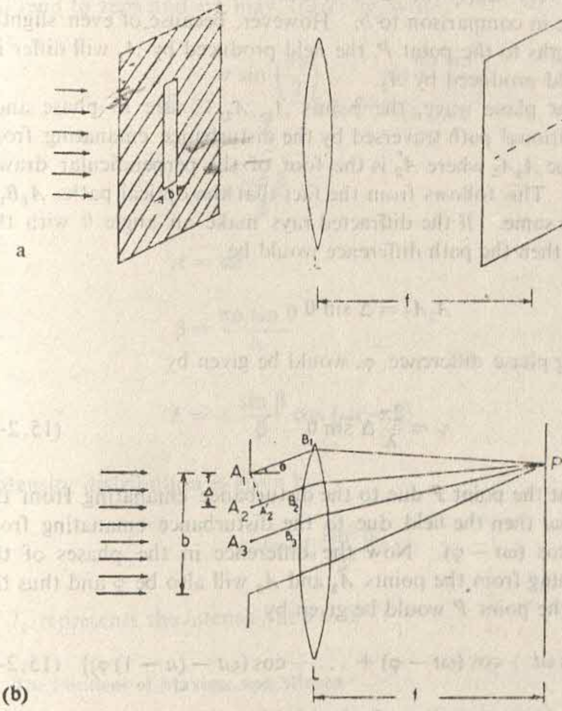


Fig. 15.3 (a) Diffraction of a plane wave incident on a long narrow slit of width b. Notice that the spreading occurs along the width of the slit. (b) In order to calculate the diffraction pattern, the slit is assumed to consist of a large number of equally spaced point sources.

15.3(b)]. Thus, if the number of point sources be n then

$$b = (n - 1)\Delta \tag{15.2-1}$$

We will now calculate the resultant field produced by these n sources at the point P; P being an arbitrary point (on the focal plane of the lens) receiving parallel rays making an angle θ with the normal to the slit [see Fig. 15.3(b)]. Since the slit actually consists of a continuous distribution of sources, we will, in the final expression, let n go to infinity and Δ go to zero such that $n\Delta$ tends to b.

Now, at the point P, the amplitudes of the disturbances reaching from A_1, A_2, . . . will be very nearly the same because the point P is at a distance which is very large in comparison to b. However, because of even slightly different path lengths to the point P, the field produced by A_1 will differ in phase from the field produced by A_2.

For an incident plane wave, the points A_1, A_2, . . . are in phase and, therefore, the additional path traversed by the disturbance emanating from the point A_2 will be A_2A_2' where A_2' is the foot of the perpendicular drawn from A_1 on A_2B_2. This follows from the fact that the optical paths A_1B_1P and $A_2'B_2P$ are the same. If the diffracted rays make an angle θ with the normal to the slit then the path difference would be

$$A_2A_2' = \Delta \sin \theta$$

The corresponding phase difference, φ, would be given by

$$\varphi = \frac{2\pi}{\lambda} \Delta \sin \theta \tag{15.2-2}$$

Thus, if the field at the point P due to the disturbance emanating from the point A_1 is $a \cos \omega t$ then the field due to the disturbance emanating from A_2 would be $a \cos (\omega t - \varphi)$. Now the difference in the phases of the disturbance reaching from the points A_2 and A_3 will also be φ and thus the resultant field at the point P would be given by

$$E = a[\cos \omega t + \cos (\omega t - \varphi) + \ldots + \cos (\omega t - (n - 1)\varphi)] \tag{15.2-3}$$

where

$$\varphi = \frac{2\pi}{\lambda} \Delta \sin \theta.$$

Now, we had shown in Sec. 10.7 that

$$\cos \omega t + \cos (\omega t - \varphi) + \ldots + \cos [\omega t - (n - 1)\varphi]$$
$$= \frac{\sin n\varphi/2}{\sin \varphi/2} \cos [\omega t - \tfrac{1}{2}(n - 1)\varphi] \tag{15.2-4}$$

Thus

$$E = E_\theta \cos [\omega t - \tfrac{1}{2}(n - 1)\varphi] \tag{15.2-5}$$

where the amplitude E_θ of the resultant field would be given by

$$E_\theta = a\,\frac{\sin (n\varphi/2)}{\sin \varphi/2} \tag{15.2-6}$$

In the limit of $n \to \infty$ and $\Delta \to 0$ in such a way that $n\Delta \to b$, we have

$$\frac{n\varphi}{2} = \frac{\pi}{\lambda}\, n\Delta \sin\theta \to \frac{\pi}{\lambda}\, b \sin\theta$$

Further

$$\varphi = \frac{2\pi}{\lambda}\Delta \sin\theta = \frac{2\pi}{\lambda}\,\frac{b \sin\theta}{n}$$

would tend to zero and we may, therefore, write

$$E_\theta \approx \frac{a \sin\left(\dfrac{n\varphi}{2}\right)}{\dfrac{\varphi}{2}} = na\,\frac{\sin \dfrac{\pi b \sin\theta}{\lambda}}{\dfrac{\pi b \sin\theta}{\lambda}}$$

$$= A\,\frac{\sin\beta}{\beta} \tag{15.2-7}$$

where

$$A = na$$

and

$$\beta = \frac{\pi b \sin\theta}{\lambda} \tag{15.2-8}$$

Thus

$$E = A\,\frac{\sin\beta}{\beta}\cos(\omega t - \beta) \tag{15.2-9}$$

The intensity distribution is given by

$$I = I_0\,\frac{\sin^2\beta}{\beta^2} \tag{15.2-10}$$

where I_0 represents the intensity at $\theta = 0$.

15.2.1 The Positions of Maxima and Minima

The variation of the intensity with β is shown in Fig. 15.4 (a). It is obvious from Eq. (15.2-9) that the intensity is zero when

$$\beta = m\pi,\ m \neq 0 \tag{15.2-11}$$

[When $\beta = 0$, $\left(\dfrac{\sin\beta}{\beta}\right) = 1$ and $I = I_0$ which corresponds to the maximum of the intensity]. Substituting the value of β one obtains

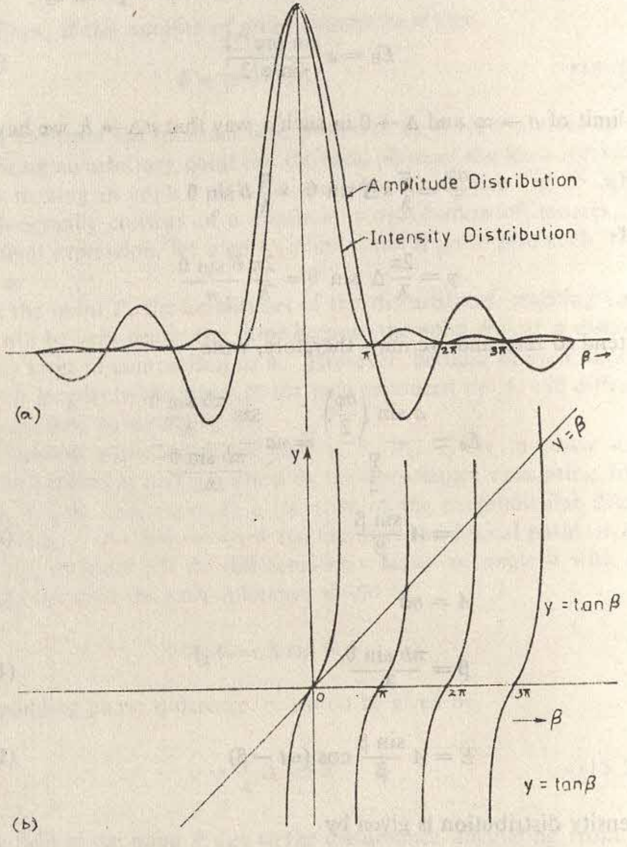


Fig. 15.4 (a) The single slit Fraunhofer diffraction pattern. (b) Graphical method for determining the roots of the equation tan β = β.

$$b \sin \theta = m\lambda;\ m = \pm 1, \pm 2, \pm 3, \ldots \quad \text{(MINIMA)} \quad (15.2\text{-}12)$$

as the conditions for minima. The first minimum occurs at $\theta = \pm \sin^{-1}\left(\frac{\lambda}{b}\right)$; the second minimum at $\theta = \pm \sin^{-1}\left(\frac{2\lambda}{b}\right)$, etc. Since $\sin \theta$ cannot exceed unity, the maximum value of m is the integer which is less than (and closest to) $\frac{b}{\lambda}$.

The positions of minima can directly be obtained by simple qualitative arguments. Let us consider the case $m = 1$. The angle θ satisfies the equation

$$b \sin \theta = \lambda \qquad (15.2\text{-}13)$$

We divide the slit into two halves as shown in Fig. 15.5. Consider two points A and A' separated by a distance $b/2$. Clearly the path difference

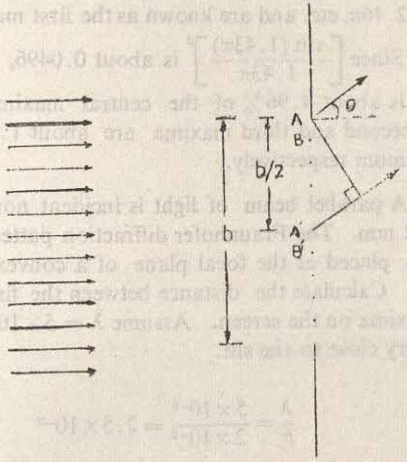


Fig. 15.5 The slit is divided into two halves for deriving the condition for the first minimum.

between the disturbances (reaching the point P) emanating from A and A' is $\frac{b}{2}\sin\theta$ which in this case is $\frac{\lambda}{2}$. The corresponding phase difference will be π and the resultant disturbance will be zero. Similarly, the disturbance from the point B will be cancelled by the disturbance reaching from the point B'. Thus the resulting disturbance due to the upper half of the slit will be cancelled by the disturbances reaching from the lower half and the resultant intensity will be zero. In a similar manner when

$$b\sin\theta = 2\lambda \tag{15.2-14}$$

we divide the slit into four parts; the first and second quarters cancelling each other and the third and fourth quarters cancelling each other. Similarly when $m = 3$, the slit is divided in six parts and so on.

In order to determine the positions of maxima, we differentiate Eq. (15.2-10) with respect to β and set it equal to zero. Thus

$$\frac{dI}{d\beta} = I_0\left[\frac{\sin\beta\cos\beta}{\beta^2} - \frac{2\sin^2\beta}{\beta^3}\right] = 0$$

or

$$\sin\beta\,[\beta - \tan\beta] = 0 \tag{15.2-15}$$

The condition $\sin\beta = 0$, or $\beta = m\pi$ correspond to minima. The conditions for maxima are roots of the following transcendental equation

$$\tan\beta = \beta \quad \text{(MAXIMA)} \tag{15.2-16}$$

The root $\beta = 0$ corresponds to the central maximum. The other roots can be found by determining the points of intersections of the curves $y = \beta$ and $y = \tan \beta$ [see Fig. 15.4(b)]. The intersections occur at $\beta = 1.43\pi$, $\beta = 2.46\pi$, etc. and are known as the first maximum, the second maximum, etc. Since $\left[\frac{\sin(1.43\pi)}{1.43\pi}\right]^2$ is about 0.0496, the intensity of the first maximum is about 4.96% of the central maximum. Similarly, the intensity of the second and third maxima are about 1.68% and 0.83% of the central maximum respectively.

Example 15.1 A parallel beam of light is incident normally on a narrow slit of width 0.2 mm. The Fraunhofer diffraction pattern is observed on a screen which is placed at the focal plane of a convex lens whose focal length is 20 cm. Calculate the distance between the first two minima and the first two maxima on the screen. Assume $\lambda = 5 \times 10^{-5}$ cm and that the lens is placed very close to the slit.

Solution.

$$\frac{\lambda}{b} = \frac{5 \times 10^{-5}}{2 \times 10^{-2}} = 2.5 \times 10^{-3}$$

Since conditions for minima are given by Eq. (15.2-11), the angles of diffraction corresponding to first and second minima would be approximately* 2.5×10^{-3} and 5×10^{-3} radians respectively. These will be separated by a distance $(5 \times 10^{-3} - 2.5 \times 10^{-3}) \times 20 = 0.05$ cm on the focal plane of the lens. Similarly, the first and second maxima occur at

$$\beta = 1.43\pi \text{ and } 2.46\pi$$

respectively. Thus

$$b \sin \theta = 1.43\lambda \text{ and } 2.46\lambda$$

or

$$\sin \theta = 1.43 \times 2.5 \times 10^{-3} \text{ and } 2.46 \times 2.5 \times 10^{-3}$$

Consequently, the maxima will be separated by distance

$$(2.46 - 1.43) \times 2.5 \times 10^{-3} \times 20 \simeq 0.05 \text{ cm}.$$

In Plate 10 we have given the actual single slit diffraction pattern (as seen on a screen) for the following values of slit widths: 8.8×10^{-3} cm, 1.76×10^{-2} cm, 3.5×10^{-2} cm and 7.0×10^{-2} cm. The light wavelength used was 6328 Å $= 6.328 \times 10^{-5}$ cm. We may note the following two points:

(i) The spreading is only in the direction of the width of the slit. This is because of the fact that the lengths of the slits were large compared to their widths.

(ii) The values of λ/b corresponding to the four slit widths are

---

*Since the angles are small we are assuming $\sin \theta \simeq \theta$.

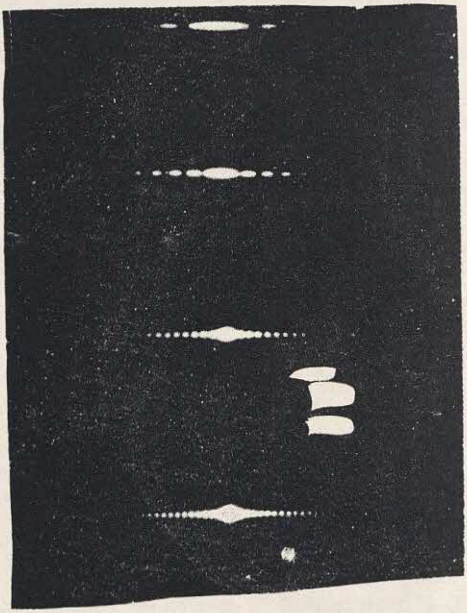
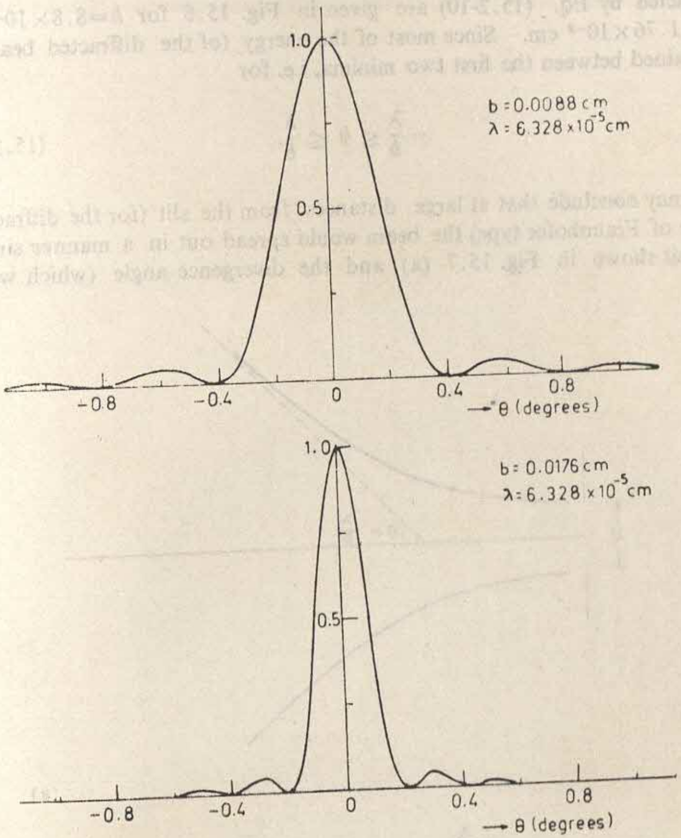


Plate 10 The single slit diffraction patterns corresponding to $b=0.0088$, 0.0176, 0.035 and 0.070 cm respectively. The wavelength of the light used is 6.328×10^{-5} cm. (After H. F. Meiners, *Physics Demonstration Experiments*, Volume II, The Ronald Press Co., New York, 1970; used with permission.)



(b)

Fig. 15.6 The intensity distributions as calculated using Eq. (15.2–10) for $b = 0.0088$ and 0.0176 cm. ($\lambda = 6.328 \times 10^{-5}$ cm).

7.191×10^{-3}, 3.595×10^{-3}, 1.808×10^{-3} and 0.904×10^{-3}. Thus the diffraction angle at which the first minimum will occur will be

$$\theta \simeq \sin\theta = 7.191 \times 10^{-3},\ 3.595 \times 10^{-3},\ 1.808 \times 10^{-3} \text{ and } 0.904 \times 10^{-3};$$

where the angles are measured in radians.* The intensity distribution as predicted by Eq. (15.2-10) are given in Fig. 15.6 for $b=8.8\times10^{-3}$ cm and 1.76×10^{-2} cm. Since most of the energy (of the diffracted beam) is contained between the first two minima, i.e. for

---

*Plate 10 corresponds to the photographic film being 15 feet away from the slit. Thus it records the Fraunhofer pattern (see Sec. 16.6) and for $b = 8.8 \times 10^{-3}$ cm, 1.76×10^{-2} cm, 3.5×10^{-2} cm and 7.0×10^{-2} cm, the first minimum occurs at a distance of 3.288 cm, 1.644 cm, 0.827 cm and 0.413 cm from the central maximum.

$$-\frac{\lambda}{b} \lesssim \theta \lesssim \frac{\lambda}{b}, \tag{15.2-17}$$

one may conclude that at large distances from the slit (for the diffraction to be of Fraunhofer type) the beam would spread out in a manner similar to that shown in Fig. 15.7 (a) and the divergence angle (which would

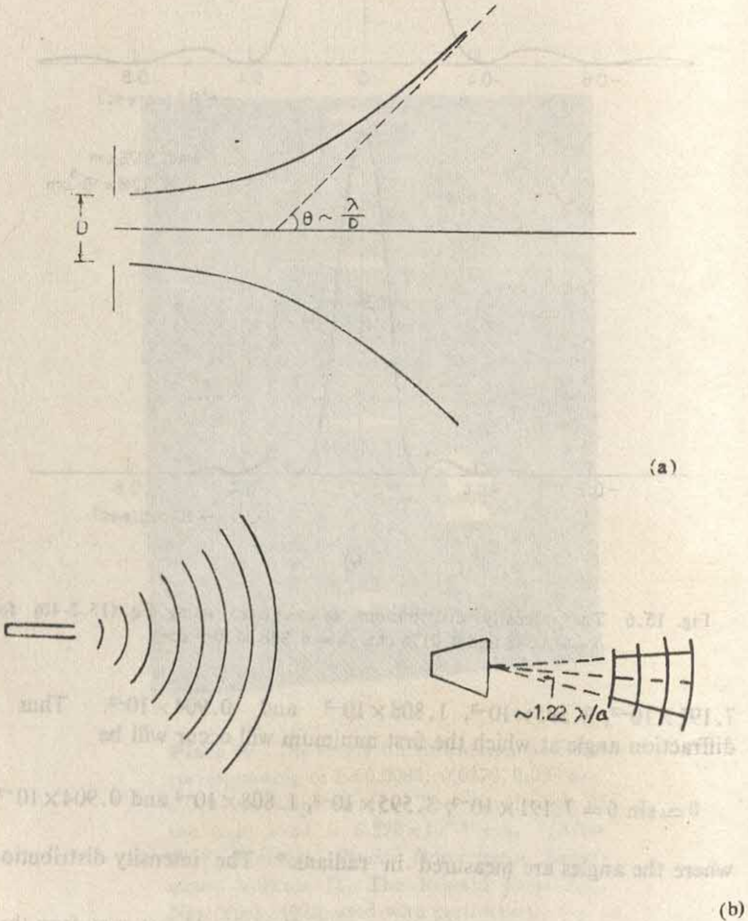


Fig. 15.7 (a) The diffraction divergence of a beam of finite circular cross-section.
(b) The directionality of sound waves increases with increase in diameter of speakers.

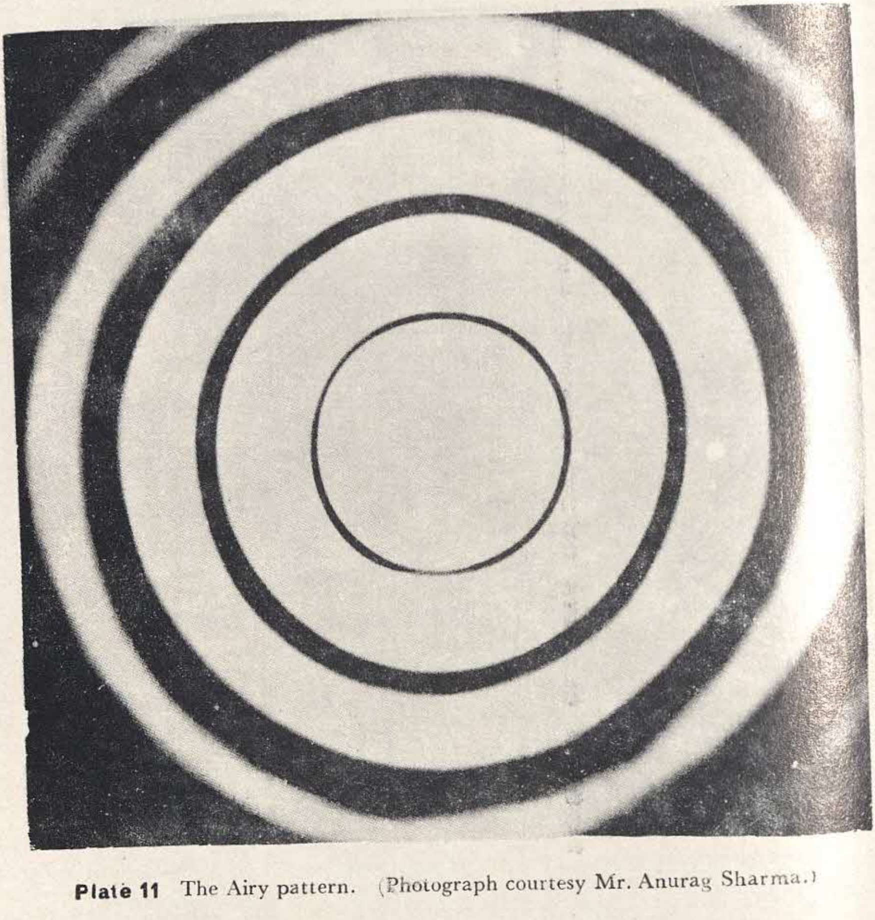


Plate 11 The Airy pattern. (Photograph courtesy Mr. Anurag Sharma.)

contain most of the energy) would be given by (see also Plate 4)

$$\Delta\theta \sim \lambda/b \qquad (15.2\text{-}18)$$

where we have assumed $b \gg \lambda$ so that $\Delta\theta$ is small.* As a simple corollary of this we may mention that the directionality of sound waves increases with increase in diameter of speakers [see Fig. 15.7(b)].

In the limit of $\lambda \rightarrow 0$, the diffraction effects will be completely absent. In fact the branch of optics in which we assume $\lambda \rightarrow 0$ is known as geometrical optics. Further, a ray is defined as the direction of propagation of energy as $\lambda \rightarrow 0$ (see Sec. 2.1).

15.3 Diffraction by a Circular Aperture

In Sec. 15.2 we have shown that when a plane wave is incident on a long narrow slit (of width b), then the emergent wave tends to spread out (along the width of the slit) with angular divergenee $\sim \lambda/b$. In a similar manner one can discuss the diffraction of a plane wave by a circular aperture (see Fig. 15.8 (a)]. Because of the rotational symmetry of the system, we can say that the diffraction pattern should be a set of concentric bright and dark rings. Indeed, the detailed mathematical analysis shows that the intensity distribution is given by (see Sec. 16.9)

$$I = I_0 \left[\frac{2J_1(v)}{v} \right]^2 \qquad (15.3\text{-}1)$$

where, $v = \frac{2\pi}{\lambda} a \sin\theta$, a being the radius of the circular aperture and θ, the angle of diffraction. The function $J_1(v)$ is known as the Bessel function of the first order. The variation of I with v is plotted in Fig. 15.8(b) and the corresponding diffraction pattern, which is called the Airy pattern is shown in Plate 11. The minima of the intensity are obtained at the zeroes of $J_1(v)$, which occur at

$$v = 3.833,\ 5.136,\ 7.016,\ \ldots \qquad (15.3\text{-}2)$$

thus, the first dark ring appears when

$$\sin\theta = \frac{3.833\lambda}{2\pi a} = \frac{1.22\lambda}{D} \qquad (15.3\text{-}3)$$

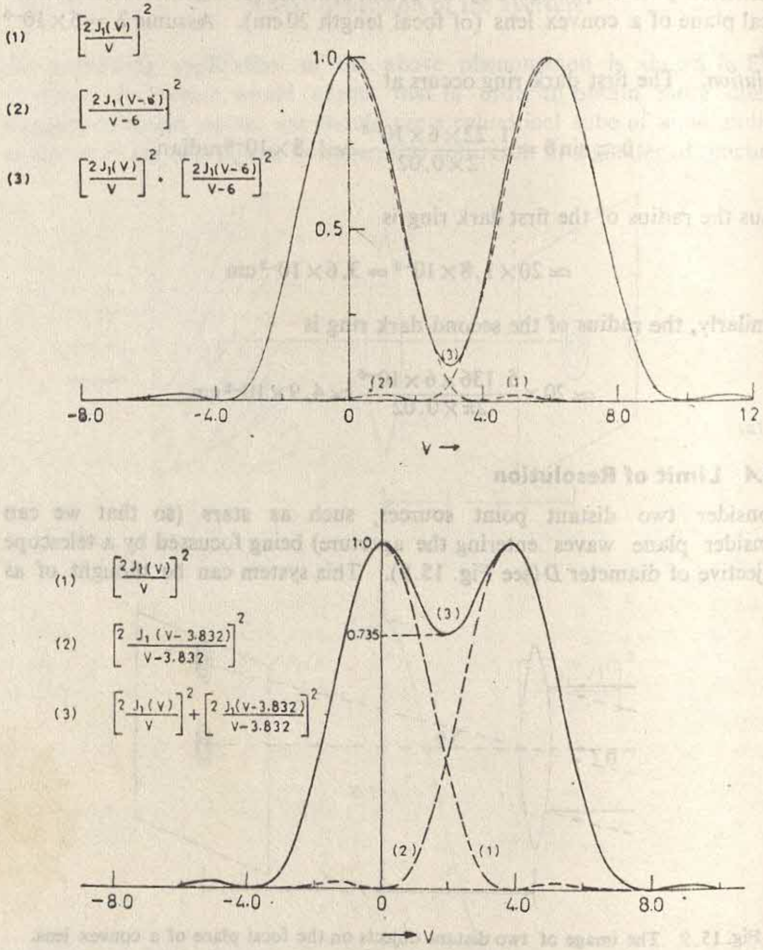
where $D(=2a)$ is the diameter of the aperture. Since about 84% of energy is contained within the first dark ring (see Fig. 16.19), one can say that the angular spreading of the beam is approximately given by (see also Plate 4).

---

*For small values of b, Eq. (15.2-18) is not valid. For example, for $b = \lambda/4$ the intensity falls off very slowly with θ and even the first minimum does not appear in the diffraction pattern; at $\theta = \pi/2$, the intensity reduces by only 20% of the value at $\theta = 0$. The light almost uniformly spreads out from the slit.

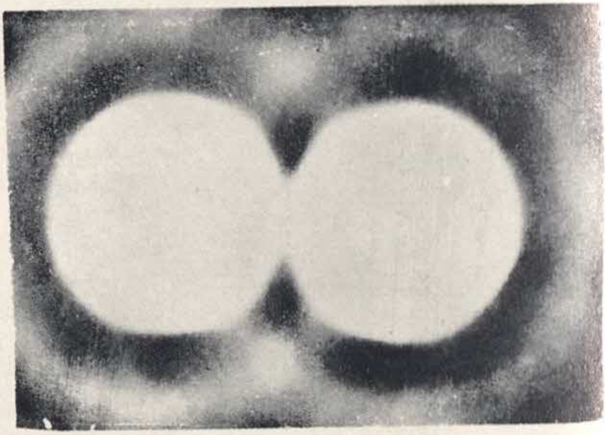
diameters of the Airy rings will be determined by the diameter of the objective, its focal length and the wavelength of light.

In Fig. 15.9 the Airy patterns are shown to be quite far away from each other and, therefore, the two objects are said to be well resolved. Since the radius of the first dark ring is $\frac{1.22\lambda}{D} f$, the Airy patterns will

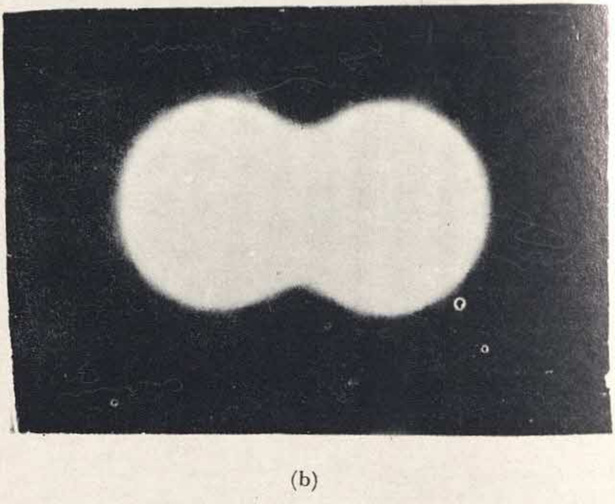


(b)

Fig. 15.10 The dashed curves correspond to the intensity distribution produced by two point sources (of equal intensity) independently; the solid curves represent the resultant. (a) and (b) correspond to the cases when the objects are well resolved and just resolved (according to the Rayleigh criterion) respectively.



(a)



(b)

Plate 12 The actual diffraction patterns corresponding to two point sources when they are (a) well resolved, (b) just resolved. (Photographs courtesy Mr. Anurag Sharma.)

overlap more for smaller values of *D* and hence for better resolution one requires a larger diameter of the objective.

In Fig. 15.10(a) we have plotted the resultant intensity distribution (produced by two distant objects corresponding to an angular separation $\simeq \frac{6}{\pi}\frac{\lambda}{D}$) and as can be seen the two images are clearly resolved. Fig. 15.11 corresponds to $\Delta\theta \approx \frac{2}{\pi}\frac{\lambda}{D}$ and the overall intensity distribution has only one peak and, therefore, the two points cannot be resolved at all. If the angular separation of the two objects is $\frac{1.22\lambda}{D}$ then the central spot of one pattern falls on the first minimum of the second and the objects are said to be just resolved. This criterion of limit of resolution is called the Rayleigh criterion of resolution and the intensity distribution corresponding to this is plotted in Fig. 15.10(b). The actual diffraction patterns are shown in Plate 12.

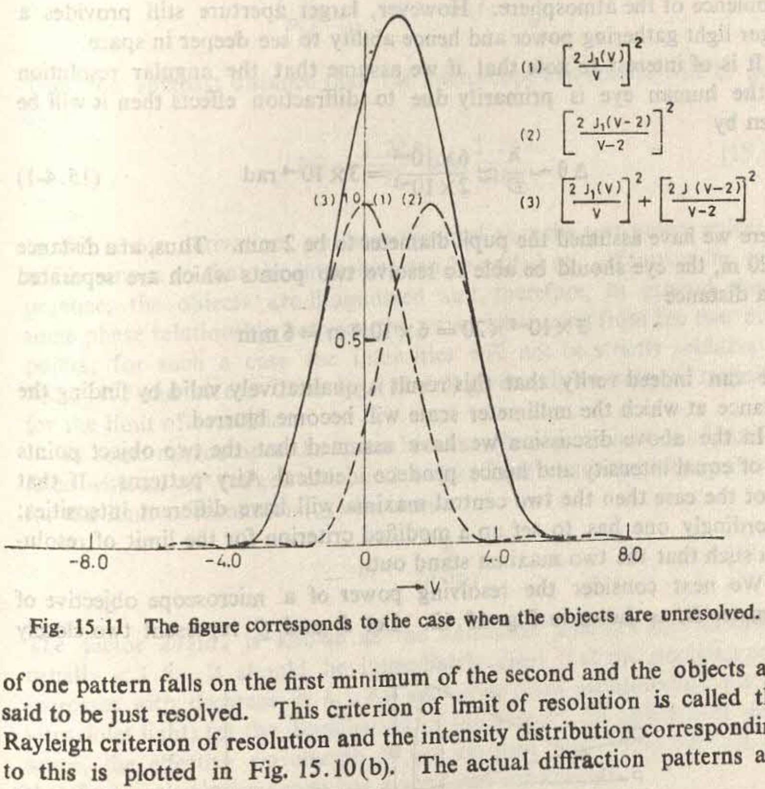


Fig. 15.11 The figure corresponds to the case when the objects are unresolved.

In order to get a numerical appreciation of the above results we consider a telescope objective whose diameter and focal length are 5 cm and 30 cm respectively. Assuming the light wavelength to be 6×10^{-5} cm one finds that the minimum angular separation of two distant objects which can be barely resolved will be

$$\frac{1.22\lambda}{D} = \frac{1.22\times6\times10^{-5}}{5} \simeq 1.5\times10^{-5} \text{ radians}$$

Further, the radius of the first dark ring (of the Airy pattern) will be

$$\frac{1.22\,\lambda}{D}\times\text{focal length}=\frac{1.22\times 6\times 10^{-5}}{5}\times 30$$

$$\simeq 4.5\times 10^{-4}\,\text{cm}.$$

It is immediately obvious that the larger the diameter of the aperture, the better will be its resolving power. For example, the diameter of the largest telescope objective is about 80″ and the corresponding angular separation of the objects that it can resolve is ∼0.07 sec. This very low limit of resolution is never achieved in ground based telescopes due to the turbulence of the atmosphere. However, larger aperture still provides a larger light gathering power and hence ability to see deeper in space.

It is of interest to note that if we assume that the angular resolution of the human eye is primarily due to diffraction effects then it will be given by

$$\Delta\theta\sim\frac{\lambda}{D}\approx\frac{6\times 10^{-5}}{2\times 10^{-1}}=3\times 10^{-4}\,\text{rad} \qquad (15.4\text{-}1)$$

where we have assumed the pupil diameter to be 2 mm. Thus, at a distance of 20 m, the eye should be able to resolve two points which are separated by a distance

$$3\times 10^{-4}\times 20=6\times 10^{-3}\,\text{m}=6\,\text{mm}$$

One can indeed verify that this result is qualitatively valid by finding the distance at which the millimeter scale will become blurred.

In the above discussion we have assumed that the two object points are of equal intensity and hence produce identical Airy patterns. If that is not the case then the two central maxima will have different intensities; accordingly one has to set up a modified criterion for the limit of resolution such that the two maxima stand out.

We next consider the resolving power of a microscope objective of diameter *D*, as shown in Fig. 15.12. Let *P* and *Q* represent two closely spaced points which are to be viewed through the microscope. The point *P* produces a spherical wavefront, which will form an Airy pattern around

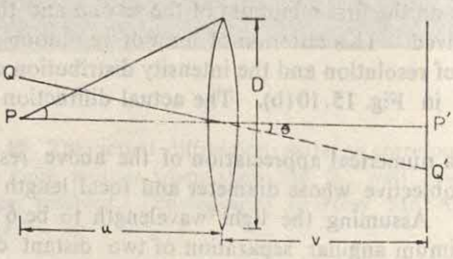


Fig. 15.12 The resolving power of a microscope objective.

the paraxial image point P'. Clearly, the radius of the first dark ring in the Airy pattern will be $\frac{1.22\lambda}{D}v$ where v represents the distance between the lens and the image plane. If Q' represents the paraxial image of the point Q then for the points P' and Q' to be just resolved we must have

$$P'Q' \approx \frac{1.22\lambda}{D}v \tag{15.4-2}$$

Further, from the geometry of the system,

$$\frac{P'Q'}{PQ} = \frac{v}{u} \quad \text{and} \quad \sin i \simeq \frac{D/2}{u}$$

Thus the smallest distance that the microscope can resolve will be given by

$$PQ = \frac{1.22\lambda}{D}u \simeq \frac{1.22\lambda}{\sin i} \tag{15.4-3}$$

In the above derivation, we have assumed that the two object points are self luminous so that the intensities can be added up. However, in actual practice, the objects are illuminated and, therefore, in general, there is some phase relationship between the waves emanating from the two object points; for such a case the intensities will not be strictly additive (see Secs. 11.3 and 14.6). Thus Eq. (15.4-3) gives only an order of magnitude for the limit of resolution.

In some microscopes the space between the object and the objective is filled with an oil* (of refractive index n) and for such a case the expression for the limit of resolution (Δ) becomes

$$\Delta \sim \frac{\lambda}{2n\sin i} \tag{15.4-4}$$

The factor $2n\sin i$ is known as the numerical aperture and its value is usually ≤ 1.6. It should be immediately clear that the resolving power increases with decrease in λ. As such, one often uses blue light (or even ultraviolet light) for the illumination of the object. In an electron microscope, the effective wavelength of the electron beam is $\sim 10^{-9}$ cm and, therefore, such a microscope has a very high resolution.

15.5 Two Slit Fraunhofer Diffraction Pattern

In Sec. 15.3 we had studied the Fraunhofer diffraction pattern produced by a slit of width b and had found that the intensity distribution consisted of maxima and minima. In this section we will study the Fraunhofer diffraction pattern produced by two parallel slits (each of width b) separated by a distance d. We would find that the resultant intensity distribution is a product of the single slit diffraction pattern and the interference pattern produced by two point sources separated by a distance d.

---

*These are known as oil immersion objectives.

In order to calculate the diffraction pattern we use a method similar to that used for the case of a single slit and assume that the slits consist of a large number of equally spaced point sources and that each point on the slit is a source of Huygens' secondarv wavelets. Let the point sources be at $A_1, A_2, A_3, \ldots$ (in the first slit) and at $B_1, B_2, B_3 \ldots$ (in the second slit) [see Fig. 15.13]. As before, we assume that the distance between two consecutive points in either of the slits is Δ. If the diffracted rays make an angle θ with the normal to the plane of the slits, then the path difference between the disturbances reaching the point P from two consecutive points in a slit will be $\Delta \sin\theta$. The field produced by the first slit at the point P will, therefore, be given by (see Eq. 15.2-9)

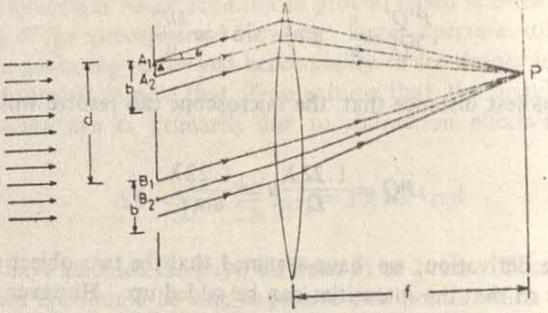


Fig. 15.13 Fraunhofer diffraction of a plane wave incident normally on a double slit.

$$E_1 = A\frac{\sin\beta}{\beta}\cos(\omega t - \beta) \tag{15.5-1}$$

Similarly, the second slit will produce a field

$$E_2 = A\frac{\sin\beta}{\beta}\cos(\omega t - \beta - \Phi_1) \tag{15.5-2}$$

at the point P, where

$$\Phi = \frac{2\pi}{\lambda} d \sin\theta \tag{15.5-3}$$

represents the phase difference between the disturbances (reaching the point P) from two corresponding points on the slits; by corresponding points we imply pairs of points like (A_1, B_1), $(A_2, B_2), \ldots$ which are separated by a distance d. Hence the resultant field will be

$$E = E_1 + E_2$$

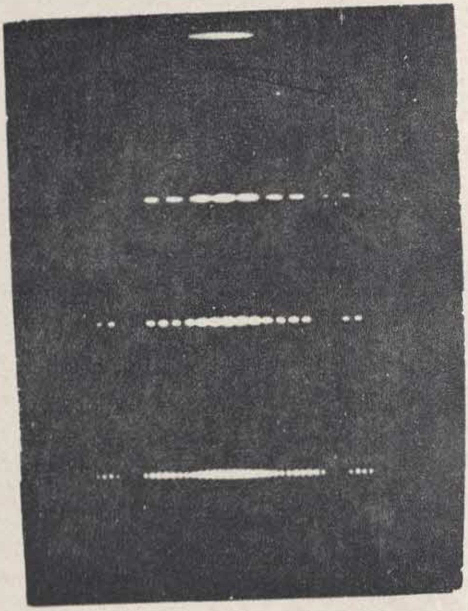


Plate 13 The double slit Fraunhofer diffraction pattern corresponding to $b=0.0088$ cm and $\lambda=6.328\times10^{-5}$ cm. The values of d are 0, 0.0176, 0.035 and 0.070 cm respectively. (After H. F. Meiners, *Physics Demonstration Experiments*, Volume II, The Ronald Press Co., New York, 1970: used with permission.)

$$= A\,\frac{\sin\beta}{\beta}\,[\cos(\omega t-\beta)+\cos(\omega t-\beta-\Phi_1)] \qquad (15.5\text{-}4)$$

which represents the interference of two waves, each of amplitude $A\frac{\sin\beta}{\beta}$ and differing in phase by Φ_1. Eq. (15.5-4) can be rewritten in the form

$$E = A\,\frac{\sin\beta}{\beta}\cos\gamma\,\cos\left(\omega t-\frac{1}{2}\beta-\frac{1}{2}\Phi_1\right) \qquad (15.5\text{-}5)$$

where

$$\gamma=\frac{\Phi_1}{2}=\frac{\pi}{\lambda}\,d\sin\theta \qquad (15.5\text{-}6)$$

The intensity distribution will be of the form

$$I = 4I_0\,\frac{\sin^2\beta}{\beta^2}\cos^2\gamma \qquad (15.5\text{-}7)$$

where $I_0\ \sin^2\beta/\beta^2$ represents the intensity distribution produced by one of the slits. As can be seen, the intensity distribution is a product of two terms; the first term $(\sin^2\beta/\beta^2)$ represents the diffraction pattern produced by a single slit of width b and the second term $(\cos^2\gamma)$ represents the interference pattern produced by two point sources separated by a distance d. Indeed, if the slit widths are very small (so that there is almost no variation of the $\sin^2\beta/\beta^2$ term with θ) then one simply obtains the Young's interference pattern (see Eq. 10.6-19).

In Plate 13 we have shown the two slit diffraction patterns corresponding to $d=0$, 0.0176, 0.035 and 0.070 cm with $b=0.0088$ cm and $\lambda=6.328\times10^{-5}$ cm. The intensity distribution as predicted by Eq. (15.5-7) is shown in Fig. 15.14.

15.5.1 Positions of Maxima and Minima

Eq. (15.5-7) tells us that the intensity is zero wherever

$$\beta=\pi,\ 2\pi,\ 3\pi\ldots \qquad (15.5\text{-}8)$$

or when,

$$\gamma=\frac{\pi}{2},\ \frac{3\pi}{2},\ \frac{5\pi}{2},\ldots \qquad (15.5\text{-}9)$$

The corresponding angles of diffraction will be given by the following equations:

$$b\sin\theta=m\lambda;\ (m=1,2,3\ldots) \qquad (15.5\text{-}10)$$

and

$$d\sin\theta=\left(n+\frac{1}{2}\right)\lambda;\ (n=0,1,2,3\ldots) \qquad 15.5\text{-}11)$$

The interference maxima occur when,

$$\gamma=0,\ \pi,\ 2\pi\ldots \qquad (15.5\text{-}12)$$

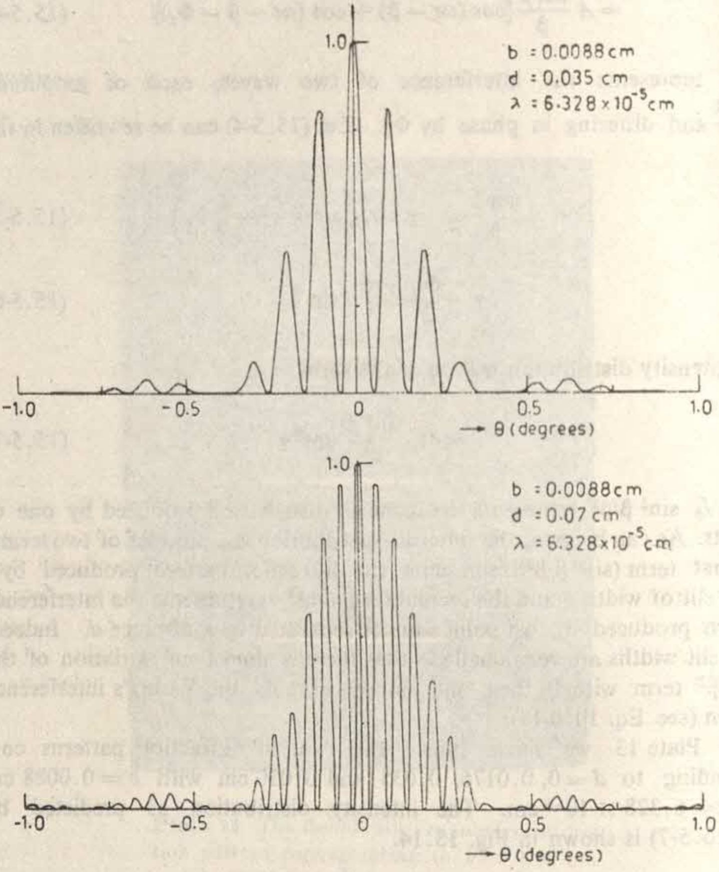


Fig. 15.14 The double slit intensity distribution as predicted by Eq. (15.5-7) corresponding to $d = 0.035$ cm and 0.070 cm respectively ($b = 0.0088$ cm, $\lambda = 6.328 \times 10^{-5}$ cm).

or when,

$$d \sin \theta = 0, \lambda, 2\lambda, 3\lambda \ldots \qquad (15.5\text{-}13)$$

The actual positions of the maxima will approximately occur at the above angles provided the variation of the diffraction term is not too rapid. Further, a maximum may not occur at all if θ corresponds to a diffraction minimum; i.e. if $b \sin \theta = \lambda, 2\lambda, 3\lambda \ldots.$ These are usually referred to as missing orders. For example, in Fig. 15.14 we can see that for $b = 0.0088$ cm, the interference maxima are extremely weak around $\theta \simeq 0.41°$; this is because of the fact that at

$$\theta \doteq \sin^{-1}\left(\frac{\lambda}{b}\right) = \sin^{-1}\left[\frac{6.328 \times 10^{-5}}{8.8 \times 10^{-3}}\right] = \sin^{-1}[7.19 \times 10^{-3}]$$

$$\simeq 0.00719 \quad \text{radians}$$

$$\simeq 0.412^\circ$$

the first minimum of the diffraction term occurs.

Example 15.3 Consider the case when $b = 8.8 \times 10^{-3}$cm, $d = 7.0 \times 10^{-2}$cm and $\lambda = 6.328 \times 10^{-5}$cm (see Fig. 15.14). How many interference minima will occur between the two diffraction minima on either side of the central maximum? In the experimental arrangement corresponding to Plate 13 the screen was placed at a distance of 15′. Calculate the fringewidth.

Solution. The interference minima will occur when Eq. (15.5-11) is satisfied; i.e. when

$$\sin\theta = \left(n + \frac{1}{2}\right)\frac{\lambda}{d} = 0.904 \times 10^{-3}\left(n + \frac{1}{2}\right); \; n = 0, 1, 2\ldots$$

$$= 0.452 \times 10^{-3},\ 1.356 \times 10^{-3},\ 2.260 \times 10^{-3},\ 3.164 \times 10^{-3},$$

$$4.068 \times 10^{-3},\ 4.972 \times 10^{-3},\ 5.876 \times 10^{-3},\ 6.780 \times 10^{-3}$$

Thus there will be sixteen minima between the two first order diffraction minima.

The angular separation between two interference maxima is approximately given by (see Eq. 15.5-13)

$$\Delta\theta \simeq \frac{\lambda}{d} = 0.904 \times 10^{-4}$$

Thus the fringewidth will be

$$15 \times 12 \times 2.54 \times 0.904 \times 10^{-4} \simeq 0.0413 \text{ cm.}$$

15.6 N-Slit Fraunhofer Diffraction Pattern

We next consider the diffraction pattern produced by N parallel slits, each of width b; the distance between two consecutive slits is assumed to be d. As before, we assume that each slit consists of n equally spaced point sources with spacing Δ (see Fig. 15.15). Thus the field at an arbitrary point P will essentially be a sum of N terms:

$$E = A\frac{\sin\beta}{\beta}\cos(\omega t - \beta) + A\frac{\sin\beta}{\beta}\cos(\omega t - \beta - \Phi_1)$$
$$+ \ldots + A\frac{\sin\beta}{\beta}\cos(\omega t - \beta - (N-1)\Phi_1) \qquad (15.6\text{-}1)$$

where the first term represents the amplitude produced by the first slit, second term by the second slit, etc. and various symbols have the same meaning as in Sec. 15.5. Rewriting Eq. (15.6-1), we get

$$E = \frac{A\sin\beta}{\beta}\left[\cos(\omega t - \beta) + \cos(\omega t - \beta + \Phi_1)\right.$$
$$\left. + \ldots + \cos(\omega t - \beta - (N-1)\Phi_1)\right]$$

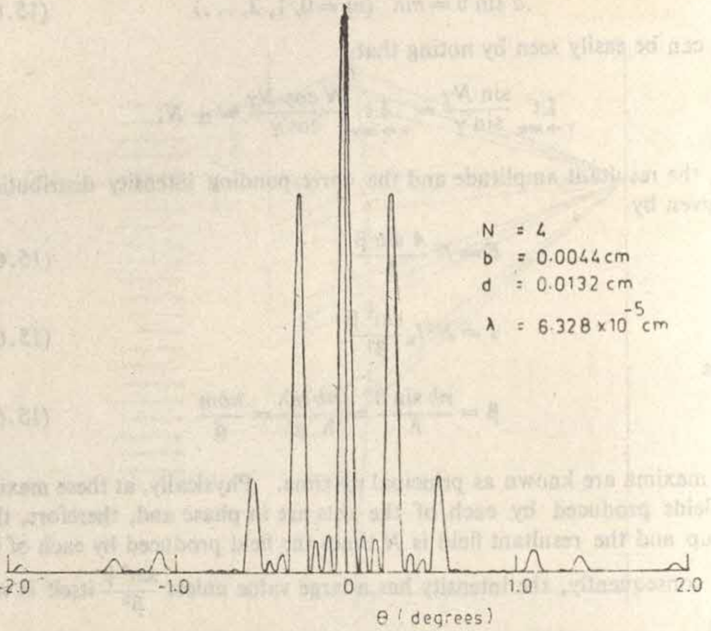


Fig. 15.16 The four slit Fraunhofer diffraction pattern as predicted by Eq. (15.6-3) corresponding to $b = 0.0044$ cm, $d = 0.0132$ cm and $\lambda = 6.328 \times 10^{-5}$ cm.

angle which also determines the minimum of the single slit diffraction pattern. This will happen when

$$d \sin \theta = m\lambda \tag{15.6-11}$$

and

$$b \sin \theta = \lambda,\ 2\lambda,\ 3\lambda,\ \ldots \tag{15.6-12}$$

are satisfied simultaneously and is usually referred to as an absent order. It should be noted that even when Eq. (15.6-12) does not hold exactly (i.e. if $b \sin \theta$ is close to an integral multiple of λ), the intensity of the corresponding principal maximum will be very weak (see, for example, Fig. 15.16 around $\theta \simeq 0.8°$).

(ii) In addition to the minima predicted by Eq. (15.6-9), we will also have the diffraction minima (see Eq. 15.6-8); however, when N is very large, the number of such minima will be very small.

15.6.2 Width of the Principal Maxima

We have shown above that in the diffraction pattern produced by N slits, the mth order principal maximum occurs at

$$d \sin \theta_m = m\lambda;\ m = 0, 1, 2, \ldots \tag{15.6-13}$$

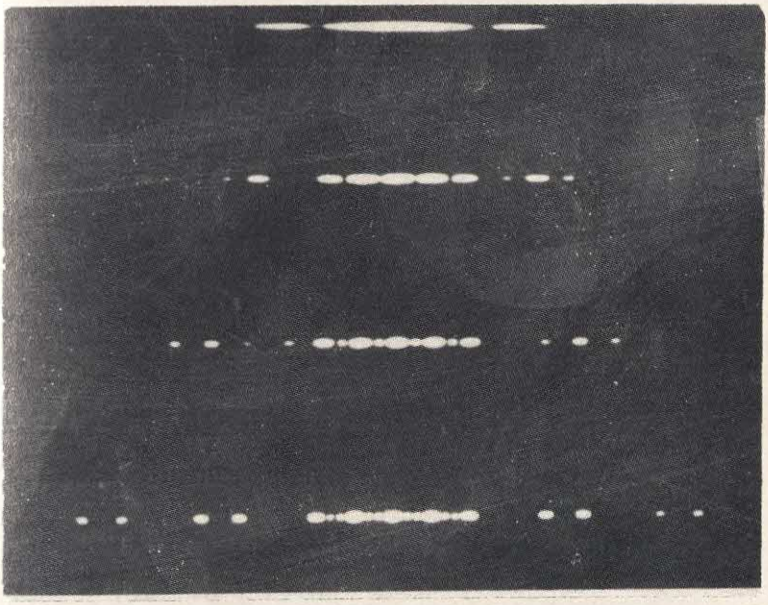


Plate 14 The multiple slit Fraunhofer diffraction patterns corresponding to $b=0.0044$ cm, $d=0.0132$ cm and $\lambda=6.328\times10^{-5}$ cm. The number of slits are 1, 2, 3 and 4 respectively. (After H.F. Meiners, *Physics Demonstration Experiments*, Volume II. The Ronald Press Co., New York, 1970; used with permission.)

Further, the minima occur at the angles given by Eq. (15.6-10). If $\theta_m + \Delta\theta_{1m}$ and $\theta_m - \Delta\theta_{2m}$ represent the angles of diffraction corresponding to the first minimum on either side of the principal maximum, then $\frac{1}{2}(\Delta\theta_{1m} + \Delta\theta_{2m})$ is known as the angular half width of the mth order principle maximum. For a large value of N, $\Delta\theta_{1m} \simeq \Delta\theta_{2m}$ which we write as $\Delta\theta_m$. Clearly,

$$d \sin(\theta_m \pm \Delta\theta_m) = m\lambda \pm \frac{\lambda}{N} \tag{15.6-14}$$

But

$$\begin{aligned} \sin(\theta_m \pm \Delta\theta_m) &= \sin\theta_m \cos\Delta\theta_m \pm \cos\theta_m \sin\Delta\theta_m \\ &\simeq \sin\theta_m \pm \Delta\theta_m \cos\theta_m \end{aligned} \tag{15.6-15}$$

Thus Eq. (15.6-14) gives us

$$\Delta\theta_m \simeq \frac{\lambda}{Nd\cos\theta_m} \tag{15.6-16}$$

which shows that the principal maximum becomes sharper as N increases.

15.7 The Diffraction Grating

In Sec. 15.6 we have discussed the diffraction pattern produced by a system of parallel equidistant slits. An arrangement which essentially consists of a large number of equidistant slits is known as a diffraction grating; the corresponding diffraction pattern is known as the grating spectrum. Since the exact positions of the principal maxima in the diffraction pattern depend on the wavelength, the principal maxima corresponding to different spectral lines (associated with a source) will correspond to different angles of diffraction. Thus the grating spectrum provides us with an easily obtainable experimental set up for determination of wavelengths. From Eq. (15.6-16) we see that for narrow principal maxima (i.e. sharper spectral lines), a large value of N is required. A good quality grating, therefore, requires a large number of slits (typically about 30000 per inch). This is achieved by ruling grooves with a diamond point on an optically transparent sheet of material; the grooves act as opaque spaces. After each groove is ruled, the machine lifts the diamond point and moves the sheet forward for the ruling of the next groove. Since the distance between two consecutive grooves is extremely small, the movement of the sheet is obtained with the help of the rotation of a screw which drives the carriage carrying it. Further, one of the important requirements of a good quality grating is that the lines should be as equally spaced as possible; consequently, the pitch of the screw must be constant, and it was not until the manufacture of a nearly perfect screw (which was achieved by Rowland in 1882) that the problem of construction of gratings was successfully solved. Rowland's arrangement gave 14438 lines per inch, corresponding to $d = 2.54/14438 = 1.693 \times 10^{-4}$ cm. For such a grating, for $\lambda = 6 \times 10^{-5}$ cm, the maximum value of m would be 2, and, therefore, only the first two orders of the spectrum will be observed. However, for $\lambda = 5 \times 10^{-5}$ cm, the third order spectrum will also be visible.

Commercial gratings are produced by taking the cast of an actual grating on a transparent film like that of cellulose acetate. An appropriate strength solution of cellulose acetate is poured on the ruled surface and allowed to dry to form a strong thin film, detachable from the parent grating. These impressions of a grating are preserved by mounting the film between two glass sheets. It may be mentioned that now a days gratings are also produced holographically, where one records the interference pattern between two plane or spherical waves. In contrast to ruled gratings, holographic gratings have a much larger number of lines/cm.

*15.7.1 The Grating Spectrum

In Sec. 15.6 we have shown that the positions of the principal maxima are given by

$$d \sin \theta = m\lambda; \quad m = 0, 1, 2, \ldots \tag{15.7-1}$$

This relation, which is also called the grating equation, can be used to study the dependence of the angle of diffraction θ on the wavelength λ. The zeroeth order principal maximum occurs at $\theta = 0$ irrespective of the wavelength. Thus, if we are using a polychromatic source (e.g. white light) then the central maximum will be of the same colour as the source itself. However, for $m \neq 0$, the angles of diffraction are different for different wavelengths and, therefore, various spectral components appear at different positions. Thus by measuring the angles of diffraction for various colours one can (knowing the value of m) determine the values of the wavelengths. It may be mentioned that the intensity is maximum for the zeroeth order spectrum (where no dispersion occurs) and it falls off as the value of m increases.

If we differentiate Eq. (15.7-1), we would obtain

$$\frac{\Delta \theta}{\Delta \lambda} = \frac{m}{d \cos \theta} \tag{15.7-2}$$

From this result we can deduce the following conclusions:

(i) Assuming θ to be very small (i.e. $\cos \theta \simeq 1$) we can see that the angle $\Delta \theta$ is directly proportional to the order of spectrum (m) for a given $\Delta \lambda$, so that for a given m, $\Delta \theta / \Delta \lambda$ is a constant. Such a spectrum is known as a normal spectrum and in this the difference in angle for two spectral lines is directly proportional to the difference in wavelengths. However, for large θ, it can easily be shown that the dispersion is greater at the red end of the spectrum.

(ii) Eq. (15.7-2) tells us that $\Delta \theta$ is inversely proportional to d, and therefore the smaller the grating element, the larger will be the angular dispersion.

Figs. 15.17 and 15.18 show schematic diagrams of the experimental arrangement for studying the grating spectrum of a polychromatic source.

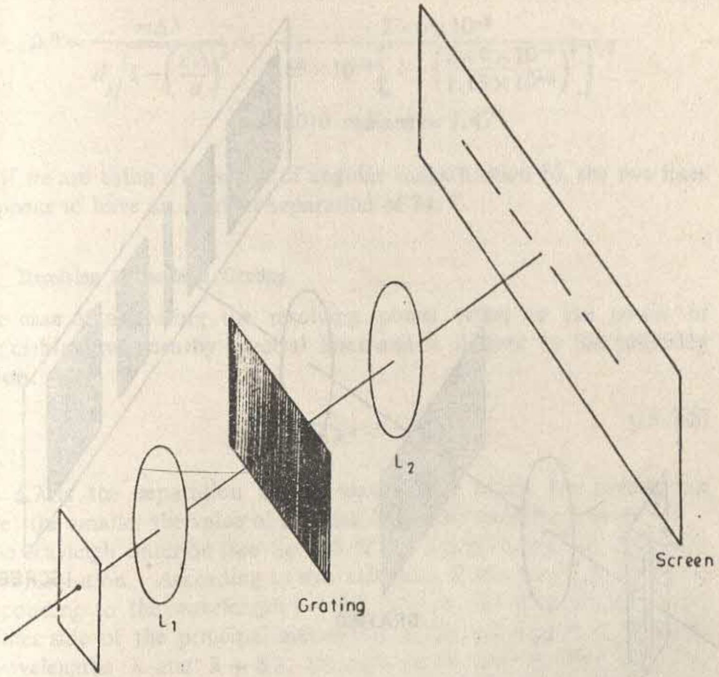


Fig. 15.17 Fraunhofer diffraction of a plane wave incident normally on a grating.

In Fig. 15.17 we have shown a small hole placed at the focal plane of the lens L_1. A parallel beam of light emerging from L_1 falls on the grating and the diffraction pattern is observed on the focal plane of the lens L_2. If instead of a hole we have a slit at the focal plane of L_1 [see Fig. (15.18)] – as it is indeed the case in a typical laboratory set up—we would have parallel beams propagating in different directions, and in the focal plane of the lens L_2 we will have a band spectrum as shown in Fig. 15.18.

The lens L_2 is the objective of a telescope and the diffraction pattern is viewed through an eyepiece. The angles of diffraction for various orders of the grating spectrum can be measured and knowing the value of d, one can calculate the wavelength of different spectral lines.

Example 15.4 Consider a diffraction grating with 15000 lines per inch. (i) Show that if we use a white light source the second and third order spectra overlap. (ii) What will be the angular separation of the D_1 and D_2 lines of sodium in the second order spectra.

Solution. (i) The grating element is

$$d=\frac{2.54}{15000}=1.69\times10^{-4}\text{ cm.}$$

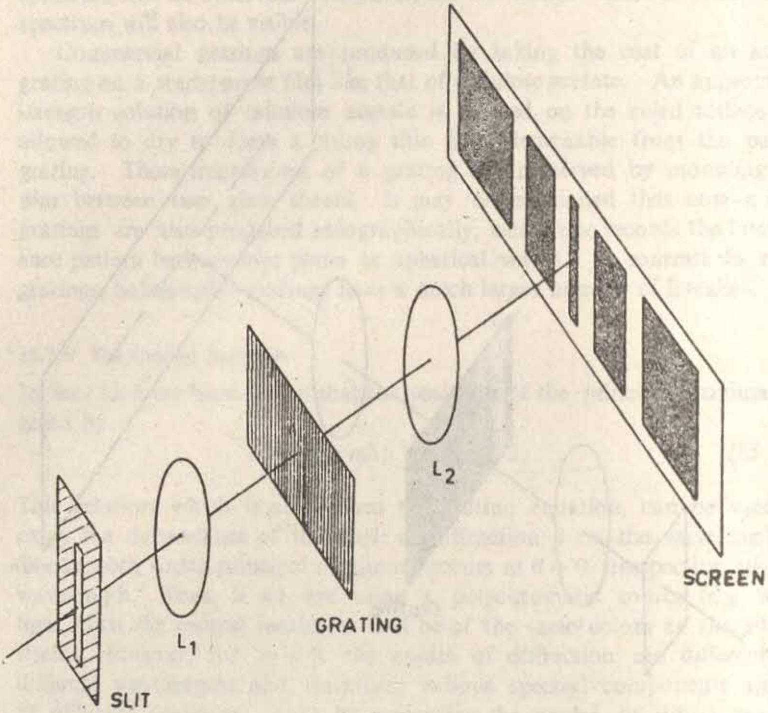


Fig. 15.18 If instead of a point source we have a slit in the focal plane of L_1 then one will obtain bands on the focal plane of L_2.

Let θ_{mv} and θ_{mr} represent the angles of diffraction for the *m*th order spectrum corresponding to the violet and red colours respectively. Thus

$$\theta_{2v} = \sin^{-1}\frac{2\times 4\times 10^{-5}}{1.69\times 10^{-4}} \simeq \sin^{-1} 0.473 \simeq 28.2^\circ$$

$$\theta_{2r} = \sin^{-1}\frac{2\times 7\times 10^{-5}}{1.69\times 10^{-4}} \simeq \sin^{-1} 0.828 \simeq 55.90^\circ$$

and

$$\theta_{3v} = \sin^{-1}\frac{3\times 4\times 10^{-5}}{1.69\times 10^{-4}} \simeq \sin^{-1} 0.710 \simeq 45.23^\circ$$

where we have assumed the wavelengths of the violet and red colours to be 4×10^{-5} cm and 7×10^{-5} cm respectively. Since $\theta_{2r} > \theta_{3v}$, the second and third order spectra will overlap. Further since $\sin\theta_{3r} > 1$; the third order spectrum for the red colour will not be observed.

(ii) Since $d \sin\theta = m\lambda$, we have for small $\Delta\lambda$:

$$(d\cos\theta)\,\Delta\theta = m(\Delta\lambda)$$

or

$$\Delta\theta = \frac{m\Delta\lambda}{d\sqrt{1-\left(\frac{m\lambda}{d}\right)^2}} \simeq \frac{2\times 6\times 10^{-8}}{1.69\times 10^{-4}\left[1-\left(\frac{2\times 6\times 10^{-5}}{1.69\times 10^{-4}}\right)^2\right]^{1/2}}$$

$$\simeq 0.0010 \text{ radians} \simeq 3.47'$$

Thus, if we are using a telescope of angular magnification 10, the two lines will appear to have an angular separation of 34.7′.

15.7.2 Resolving Power of a Grating

In the case of a grating the resolving power refers to the power of distinguishing two nearby spectral lines and is defined by the following equation:

$$R = \frac{\lambda}{\Delta\lambda}, \tag{15.7-3}$$

where $\Delta\lambda$ is the separation of two wavelengths which the grating can resolve; the smaller the value of $\Delta\lambda$, the larger the resolving power.

The Rayleigh criterion (see Sec. 15.4) can again be used to define the limit of resolution. According to this criterion, if the principal maximum corresponding to the wavelength $\lambda + \Delta\lambda$ falls on the first minimum (on the either side of the principal maximum) of the wavelength λ, then the two wavelengths λ and $\lambda + \Delta\lambda$ are said to be just resolved (see Fig. 15.19). If this common diffraction angle is represented by θ and if we are looking at the *m*th order spectrum, then the two wavelengths λ and $\lambda + \Delta\lambda$ will be just resolved if the following two equations are simultaneously satisfied:

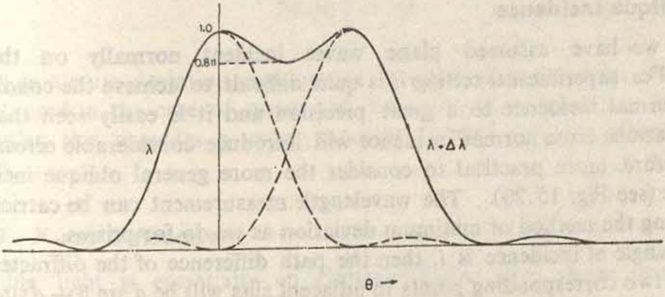


Fig. 15.19 The Rayleigh criterion for the resolution of two spectral lines.

$$d \sin\theta = m(\lambda + \Delta\lambda) \tag{15.7-4}$$

and

$$d \sin\theta = m\lambda + \frac{\lambda}{N} \tag{15.7-5}$$

Thus

$$R = \frac{\lambda}{\Delta\lambda} = mN \tag{15.7-6}$$

which implies that the resolving power depends on the total number of lines in the grating; obviously only those lines which are exposed to the incident beam. Further, the resolving power is proportional to the order of the spectrum. Thus to resolve the D_1 and D_2 lines of sodium ($\Delta\lambda = 6$ Å) in the first order, N must be at least $\frac{6\times10^{-5}}{6\times10^{-8}} \approx 1000$.

From Eq. (15.7-6) it appears that the resolving power would increase if the value of N is increased; this is, however, not quite true because the resolving power is essentially determined by the width of the grating. This can easily be seen by substituting the value of m from Eq. (15.7-1) in Eq. (15.7-6).

$$R = \frac{Nd \sin \theta_m}{\lambda} = \frac{D \sin \theta_m}{\lambda} \tag{15.7-7}$$

where $D = Nd$ represents the width of the grating.

In a similar manner one can show that the chromatic resolving power of a prism is given by

$$\frac{\lambda}{\Delta\lambda} = t\frac{dn}{d\lambda} \tag{15.7-8}$$

where t represents the length of the base of the prism and n represents the refractive index of the material of the prism.

15.8 Oblique Incidence

Till now we have assumed plane waves incident normally on the grating. For experimental setting it is quite difficult to achieve the condition of normal incidence to a great precision and it is easily seen that slight deviations from normal incidence will introduce considerable errors. It is, therefore, more practical to consider the more general oblique incidence case (see Fig. 15.20). The wavelength measurement can be carried out by using the method of minimum deviation as we do for prisms.

If the angle of incidence is i, then the path difference of the diffracted rays from two corresponding points in adjacent slits will be $d \sin \theta + d \sin i$ (see Fig. 15.20). Thus, principal maxima will occur when

$$d(\sin \theta + \sin i) = m\lambda \tag{15.8-1}$$

or

$$d[\sin(\delta - i) + \sin i] = m\lambda$$

where $\delta = i + \theta$ is the angle of deviation. For δ to be minimum we must have

$$\frac{d}{di}[\sin(\delta - i) + \sin i] = 0$$

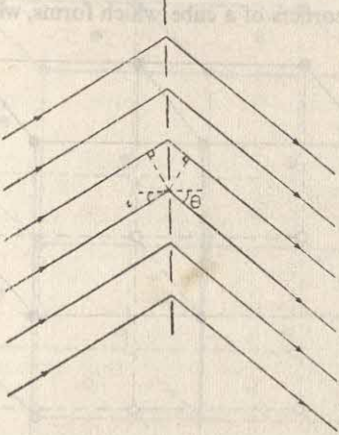


Fig. 15.20 Diffraction of a plane wave incident obliquely on a grating.

$$-\cos(\delta - i) + \cos i = 0$$

i.e.,

$$i = \delta - i = \theta$$

or,

$$i = \frac{\delta}{2} = \theta \tag{15.8-2}$$

Hence, at the position of minimum deviation, the grating condition becomes

$$2d \sin \frac{\delta}{2} = m\lambda \tag{15.8-3}$$

The minimum deviation position can be obtained in a manner similar to that used in the case of prism and since the adjustments are relatively simpler, this provides a more accurate method for the determination of λ.

15.9 X-ray Diffraction

Visible light is essentially an electromagnetic wave whose wavelength lies between 4000 Å and 7000 Å. X-rays are also electromagnetic waves whose wavelengths are ~ 1 Å. Obviously, it is extremely difficult to make slits which are narrow enough for the study of X-ray diffraction patterns. Since the interatomic spacings in a crystal are usually of the order of Angstroms, one can use it as a 3-dimensional diffraction grating for studying the diffraction of X-rays. Indeed, X-rays have extensively been used to study crystal structures.

An ideal crystal consists of a periodic arrangement of atoms (or groups of atoms) and is classified according to the geometry of this arrangement.

Fig. 15.21 (a) shows a simple cubic structure (abbreviated as SC) in which the atoms are at the corners of a cube which forms, what is known as a unit cell. The crystal is built up by the repetition of this unit cell in three dimensions. In addition, if there is an atom at the centre of each cube [see Fig. 15.21 (b)] the arrangement is known as a body centred cubic (bcc) structure. On the other hand, if instead of having an atom at the centre of the cube there is an atom at the centre of each of the six faces of the cube [see Fig. 15.22] we will have the face centred cubic (fcc) structure. Copper, silver and gold crystallize in the fcc form with the lattice parameter $a = 3.61$ Å, 4.09 Å, and 4.08 Å respectively. Metals like

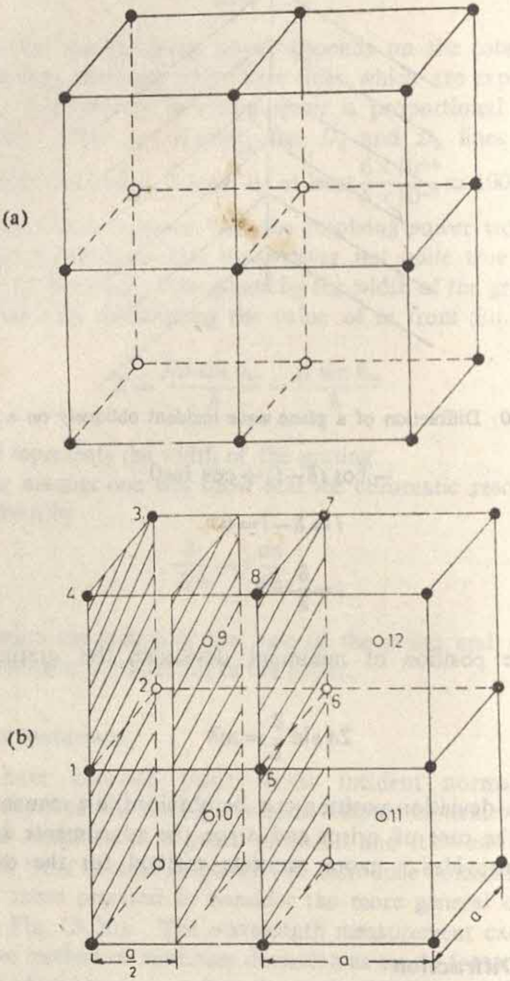


Fig. 15.21 (a) A simple cubic lattice. (b) A body centred cubic lattice.

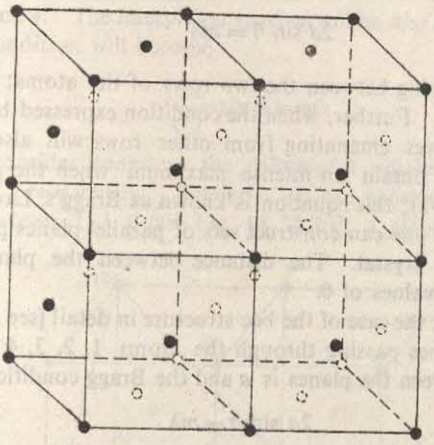


Fig. 15.22 A face centred cubic lattice.

sodium, barium, tungsten crystallize in the bcc form with $a = 4.29$ Å, 5.03 Å and 3.16 Å respectively.*

Let us consider an X-ray beam incident on a square array of atoms as shown in Fig. 15.23. W_1W_2 and W_3W_4 repsesent the incident and reflected

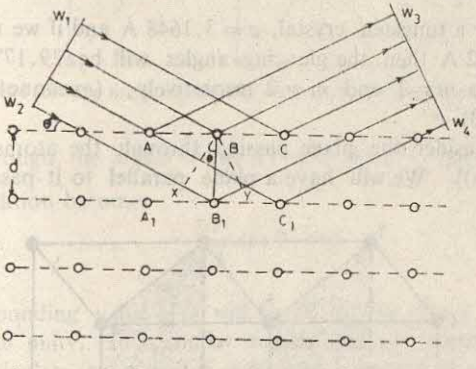


Fig. 15.23 A plane wave incident on a square array of atoms. When ($2d \sin\theta = m\lambda$), the waves scattered from different rows will be in phase.

wavefronts respectively. Obviously, the secondary wavelets emanating from the points A, B and C are in phase on W_3 W_4 (see Sec. 9.4 and Fig. 9.7); and, the waves emanating from the points A_1, B_1 and C_1 will also be in phase on W_3 W_4 if

$$XB_1 + B_1Y = m\lambda; \qquad m = 1, 2, 3, \ldots$$

*Crystal structures other than cubic are also common; e.g. zinc crystallizes into a hexagonal structure and carbon forms a diamond structure. However, the most important fact is that in all these structures there is a definite periodicity of atoms.

or, when

$$2d \sin \theta = m\lambda \tag{15.9-1}$$

where d is the spacing between the two rows of the atoms; θ is known as the glancing angle. Further, when the condition expressed by Eq. (15.9-1) is satisfied, the waves emanating from other rows will also be in phase. Thus, one should obtain an intense maximum when the glancing angle satisfies Eq. (15.9-1); this equation is known as Bragg's Law. In a three-dimensional lattice one can construct sets of parallel planes passing through the atoms of the crystal. The distance between the planes will determine the allowed values of θ.

Let us consider the case of the bcc structure in detail [see Fig. 15.21 (b)]. Consider the planes passing through the atoms 1, 2, 3, 4 and 5, 6, 7, 8. The distance between the planes is a and the Bragg condition becomes

$$2a \sin \theta = m\lambda$$

However, we must also consider the plane passing through the atoms 9 and 10 which will be midway between the two planes considered above [see Fig. 15.21 (b)] and therefore, the interplanar spacing should be taken as $a/2$ and not a; consequently, the Bragg condition becomes

$$a \sin \theta = m\lambda \tag{15.9-2}$$

For example, for a tungsten crystal, $a = 3.1648$ Å and if we use X-rays of wavelength 1.542 Å then the glancing angles will be 29.17° and 77.03° corresponding to $m = 1$ and $m = 2$ respectively; (m cannot be equal to or greater than 3).

We next consider the plane passing through the atoms 2, 3, 9, 8, 5 [see Fig. 15.24(a)]. We will have a plane parallel to it passing through

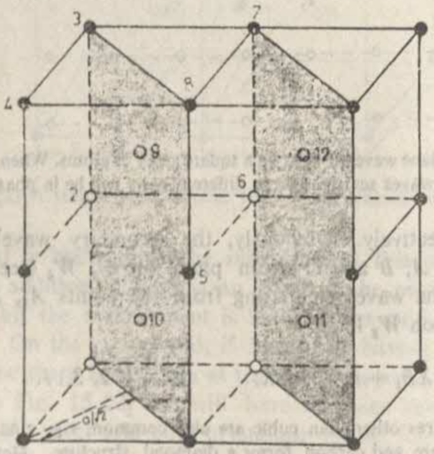


Fig. 15.24 (a) The planes in a bcc lattice separated by a distance $a/\sqrt{2}$.

the atoms 1 and 4. The interplanar spacing will be $a/\sqrt{2}$ [see Fig. 15.24(a)]. The Bragg condition will become

$$\frac{2a}{\sqrt{2}} \sin \theta = m\lambda \tag{15.9-3}$$

If we again consider tungsten, the values of θ will be 20.15° and 43.56° for $m = 1$ and 2 respectively. Similarly, we can consider the plane passing through the atoms 1, 3, 6 [see Fig. 15.24(b)]; a plane parallel to it and

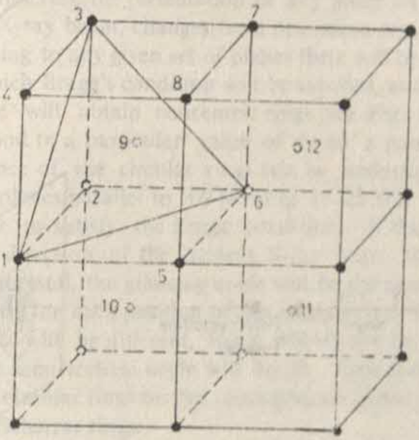


Fig. 15.24(b) The figure shows the plane passing through atoms 1, 3 and 5.

passing through the atoms 9 and 11 will be at a distance $\frac{a}{2\sqrt{3}}$ and the Bragg condition becomes

$$\frac{a}{\sqrt{3}} \sin \theta = m\lambda \tag{15.9-4}$$

The corresponding value of θ will be 57.55° for $m = 1$ and m cannot be greater than unity. In a similar manner one can construct other planes also and calculate the angles of diffraction. Indeed, it can be shown that for $\lambda = 1.542$ Å, the tungsten crystal will have only three more angles which would satisfy Bragg's Law. It is left as an exercise to the reader to show that these angles are 36.64°, 50.39° and 65.74° (for other planes d will be less than λ).

The intensity of the diffracted wave depends on the atoms per unit area in the plane under consideration. For example, corresponding to Eqs. (15.9-2) and (15.9-3) there will be one atom and two atoms in an area a^2 respectively. Thus the intensity associated with the diffracted wave corresponding to Eq. (15.9-3) will be much more than the one corresponding to Eq. (15.9-2).

From the above analysis one can see that if a monochromatic X-ray

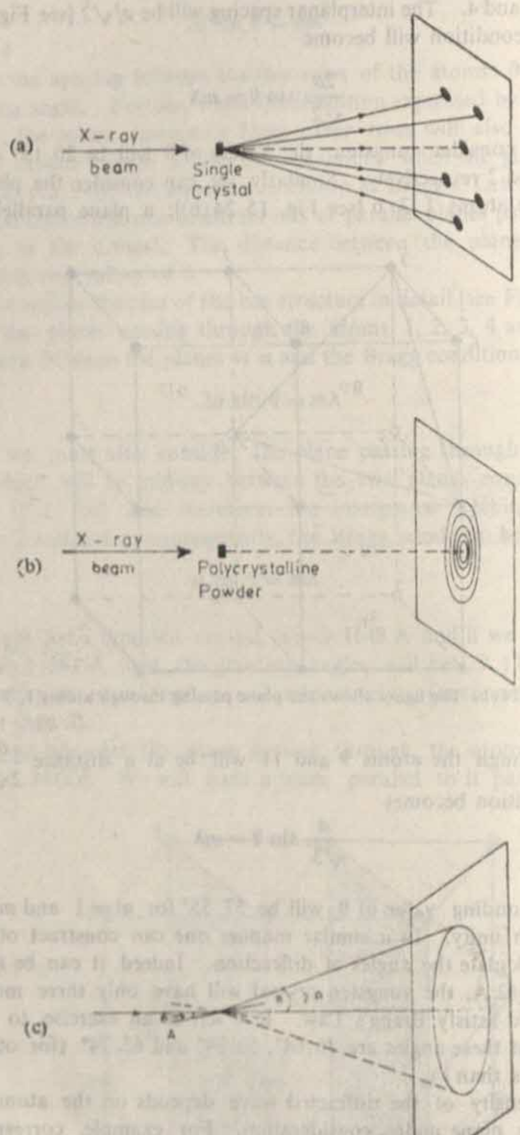


Fig. 15.25 (a) When a monochromatic X-ray beam falls on a single crystal one obtains Laue spots corresponding to the planes which satisfy Bragg's Law. (b) When a monochromatic X-ray beam falls on a polycrystalline sample one obtains the Debye-Scherrer rings. (c) Diffraction from a polycrystalline sample.

beam is allowed to fall on a single crystal, then one will obtain spots on the photographic plate placed beyond the crystal [see Fig. 15.25(a)], each spot will correspond to a definite value of θ; these spots are known as Laue spots. By a careful study of the position and intensities of these spots, the atomic arrangements in the crystal can be deduced.

Instead of a single crystal one often uses a polycrystalline material; this is known as the powder method. A polycrystalline material consists of a large number of randomly oriented micro-crystals; each micro-crystal is essentially a single crystal. As the X-ray beam passes through such a polycrystalline material, the orientation of any given set of planes, with reference to the X-ray beam, changes from one micro-crystal to the other. Thus, corresponding to any given set of planes there will be a large number of crystals for which Bragg's condition will be satisfied, and on the photographic plate one will obtain concentric rings [see Fig. 15.25(b)]; each ring will correspond to a particular value of d and a particular value of m. The appearance of the circular rings can be understood as follows: consider a set of planes parallel to AB [see Fig. 15.25 (c)]. The glancing angle θ is assumed to satisfy the Bragg condition. If the microcrystal is rotated about the direction of the incident X-ray beam, then for all positions of the microcrystal, the glancing angle will be the same for these sets of planes. Further, for each position of the microcrystal, the direction of the diffracted beam will be different, but it will always lie on the surface of the cone whose semi-vertical angle will be 2θ. Consequently, one will obtain concentric circular rings on the photographic plate; these rings are known as Debye-Scherrer rings.

While using the powder method, the photographic film is put in a cylindrical form surrounding the polycrystalline sample as shown in Fig. 15.26. Each Debye-Scherrer ring will produce an arc on the film, and when the film is unrolled, one obtains a pattern as shown in Fig. 15.27. From the position of these arcs one can calculate θ and thus obtain the interplanar spacing. From a study of the interplanar spacings one can determine the crystal structure.

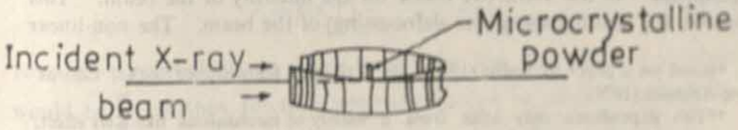
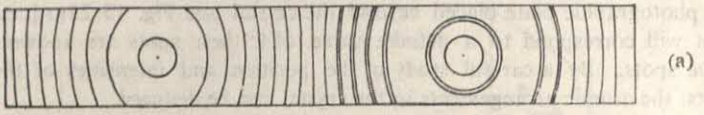
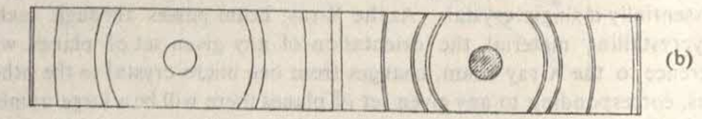


Fig. 15.26 While using the powder method the photographic film is kept in a cylindrical form as shown in the figure.

Fig. 15.27 (a) and (b) correspond to the diffraction patterns produced by sodium and copper in polycrystalline form. For example, for sodium (which is a bcc structure with $a = 4.2906$ Å the various values of θ are (14.70°, 14.74°), (21.03°, 21.09°), (26.08°, 26.15°), (30.50°, 30.59°), (34.58°, 34.68°), (38.44°, 38.56°), (42.18°, 42.32°), (45.88°, 46.03°), (49.59°, 49.76°), (53.38°, 53.58°), (57.33°, 57.56°), (61.54°, 61.82°), (66.22°,



Sodium



Copper

Fig. 15.27 Schematic diffraction patterns for (a) sodium and (b) copper.

66.56°), (79.41°, 80.23°); the two values inside the brackets correspond to the two wavelengths 1.540 Å and 1.544 Å (these are known as Cu K_{α_1} and Cu K_{α_2} lines). Because of the presence of two wavelengths one obtains double lines for each family of planes which become resolvable only at higher scattering angles.

15.10 The Self-Focussing Phenomenon*

With the availability of intense laser beams, a large number of interesting non-linear optical phenomena have been investigated. One such nonlinear phenomenon is the effect on the propagation of a light beam due to the dependence** of the refractive index on the intensity of the beam. This leads to the self-focussing (or defocussing) of the beam. The non-linear dependence of the refractive index on electric field can be assumed to be of the form

---

*Based on a paper by Sodha (1973); for a rigorous account, see Sodha, Ghatak and Tripathi (1974).

**This dependence may arise from a variety of mechanisms like Kerr effect, electrostriction, thermal effect, etc. The simplest to understand is the thermal effect which is due to the fact that when an intense optical beam having a transverse distribution of intensity propagates through an absorbing medium, a temperature gradient is set up. For example, if the beam has a Gaussian transverse intensity variation, (i.e. of the form $\exp(-r^2/a^2)$; the direction of propagation being along the z-axis), then the temperature will be maximum on the axis (i.e. $r = 0$) and will decrease with increase in the value of r. If $dn/dT > 0$, the refractive index will be maximum on the axis and the beam will undergo focussing; on the other hand if $dn/dT < 0$ the beam will undergo defocussing.

The Kerr effect arises due to the anisotropic polarizability of liquid molecules (like CS_2). An intense light wave will tend to orient the anisotropically polarized molecules such that the direction of maximum polarizability is along the direction of the electric vector; this changes the dielectric constant of the medium. On the other hand, electrostriction (which is important in solids) is the force which a non-uniform electric field exerts on a material medium; this force affects the density of

(*Contd.*)

$$n = n_0 + \tfrac{1}{2} n_2 E_0^2 \tag{15.10-1}$$

where n_0 is the refractive index of the medium in the absence of any electromagnetic field, n_2 is the non-linear constant and E_0 the amplitude of the electric field; i.e. the actual electric field is of the form

$$E = E_0 \cos (kz - \omega t) \tag{15.10-2}$$

The phenomenon of self-focussing can be physically understood by considering the propagation of a parallel cylinderical beam of radius a (see Fig. 15.28). Clearly, the refractive index in the non-illuminated and illuminated regions will be n_0 and $n_0 + \frac{1}{2}n_2 E_0^2$ respectively. Thus there will be a discontinuity in the refractive index at $r = a$. We assume n_2 to be positive and we consider the incidence of a ray at the refractive index discontinuity as shown in Fig. 15.28. Clearly, for total internal reflection to occur the angle of incidence should be greater than the critical angle, i_c, which is given by

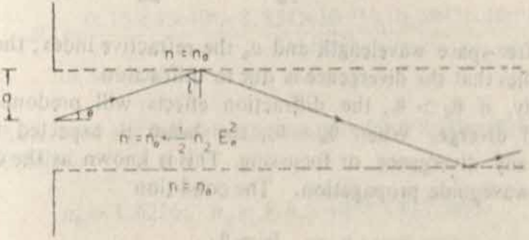


Fig. 15.28 The physical interpretation of the self-focussing phenomenon.

$$i_c = \sin^{-1}\left[\frac{n_0}{n_0 + \frac{1}{2}n_2 E_0^2}\right] \tag{15.10-3}$$

Now, for a ray making an angle θ with the axis, the angle of incidence would be $\frac{\pi}{2} - \theta$; thus, for total internal reflection to occur we must have

$$\frac{\pi}{2} - \theta > i_c = \sin^{-1}\left[\frac{n_0}{n_0 + \frac{1}{2}n_2 E_0^2}\right] \tag{15.10-4}$$

or

$$\cos\theta > \sin i_c = \cos\theta_c \tag{15.10-5}$$

where

$$\cos\theta_c \equiv \frac{n_0}{n_0 + \frac{1}{2}n_2 E_0^2} = \left[1 + \frac{1}{2}\frac{n_2}{n_0} E_0^2\right]^{1} \tag{15.10-6}$$

Since, in most cases, the nonlinearity is very weak, we may assume

---

their material, which in turn affects the refractive index. Thus, a beam having non-uniform intensity distribution along its wavefront will give rise to a refractive index variation leading to focussing (or defocussing) of the beam. For a detailed discussion on electrostriction and Kerr effect the reader is referred to Panofsky and Phillips (1962), Wagner, Haus and Marburger (1968), Sodha, Ghatak and Tripathi (1974).

$\frac{1}{2}\frac{n_2}{n_0}E_0^2 \ll 1$; consequently, θ_c will be very small and Eq. (15.10-6) takes the following approximate form:

$$1-\frac{\theta_c^2}{2}\approx 1-\frac{n_2}{2n_0}E_0^2 \tag{15.10-7}$$

Now, in the absence of any nonlinear effects, the beam will spread out due to diffraction and a very large fraction of the energy will be carried by rays making an angle less than θ_d (see Fig. 15.7) where*

$$\theta_d \simeq \frac{1}{2}\,\frac{1.22\,(\lambda_0/n_0)}{2a}=\frac{0.61\,(\lambda_0/n_0)}{2a} \tag{15.10-8}$$

λ_0 is the free-space wavelength and n_0 the refractive index; the subscript d on θ implies that the divergence is due to diffraction.

Clearly, if $\theta_d > \theta_c$, the diffraction effects will predominate and the beam will diverge. When $\theta_d \simeq \theta_c$, the beam is expected to propagate without any divergence or focussing. This is known as the condition for uniform waveguide propagation. The condition

$$\theta_c \simeq \theta_d \tag{15.10-9}$$

implies

$$\left[\frac{n_2 E_0^2}{n_0}\right]^{1/2} \simeq \frac{0.61}{2a}\,\frac{\lambda_0}{n_0}$$

or

$$E_0^2=\frac{1}{n_2 n_0}\left[\frac{0.61}{2a}\,\lambda_0\right]^2 \tag{15.10-10}$$

Now, the energy per unit volume associated with the electromagnetic wave is (see Sec. 19.5)**

$$\tfrac{1}{2}n_0^2\,\epsilon_0\,E_0^2 \tag{15.10-11}$$

where ϵ_0 is the permittivity of free space. Thus, the power of the beam is given by

P = (Area of cross-section of the beam) × (velocity) × (energy/unit volume)

or

$$P=\pi a^2\times\frac{c}{n_0}\times\tfrac{1}{2}n_0^2\,\epsilon_0\,E_0^2 \tag{15.10-12}$$

---

*The angle $2\theta_d\left(=\frac{1.22\lambda_0/n_0}{2a}\right)$ corresponds to the first minimum of the Fraunhofer diffraction pattern produced by a circular aperture of radius a (see Sec. 15.3). See also Fig. 16.18.

**Eq. (19.5-5) reads

$$\langle u\rangle=\tfrac{1}{2}\,\epsilon\,E_0^2=\tfrac{1}{2}n^2\,\epsilon_0\,E_0^2$$

where ϵ/ϵ_0 represents the dielectric constant ($=n^2$).

If we substitute the expression for E_0^2 from Eq. (15.10-10) we will get the critical power

$$P_{cr} = (0.61)^2 \frac{\pi c}{8} \frac{\epsilon_0 \lambda_0^2}{n_2}$$

$$= 0.15 \frac{c\epsilon_0 \lambda_0^2}{n_2} \qquad (15.10\text{-}13)$$

Garmire, Chiao and Townes (1966) carried out experiments on the self-focussing of a ruby laser beam ($\lambda_0 = 0.6943\ \mu m$) in CS_2 and found that the critical power was 25 ± 5 kW. On the other hand, if we use Eq. (15.10-13) we would obtain

$$P_{cr} = \frac{0.15 \times 3 \times 10^8 \times 8.854 \times 10^{-12} \times (0.6943 \times 10^{-6})^2}{2 \times 10^{-11}}$$

$$\simeq 100 \text{ kW} \qquad (15.10\text{-}14)$$

where we have used the following parameters for CS_2 [Garmire, Chiao and Townes (1966)],

$$n_0 = 1.6276; \quad n_2 = 1.8 \times 10^{-10} \text{ CGS units}$$

$$= 2 \times 10^{-19} \text{ MKS units}^* \qquad (15.10\text{-}15)$$

The MKS unit for n_2 would be (meter/volt)^2; further, the nonlinear coefficient n_2 in CS_2 is predominantly due to Kerr effect.

It should be mentioned that the result expressed by Eq. (15.10-13) is through qualitative arguments and is therefore, expected to give results which are correct within a factor of about 10. If we use the more rigorous theory (see, for example, Sodha, Ghatak and Tripathi, 1974), we would obtain 25 kW for the critical power, which compares well with the experimental data.

We may conclude that,

(i) when $P < P_{cr}$, $\theta_d > \theta_c$ and the beam will diverge due to diffraction.

(ii) when $P = P_{cr}$, $\theta_d = \theta_c$ and the beam will propagate without divergence or convergence. This is the condition for uniform waveguide propagation.

(iii) when $P > P_{cr}$, we may extrapolate that the beam will undergo focussing, which is indeed borne out by more rigorous analysis. This is known as self-focussing of the beam.

In order to have an alternative figurative description, we note that if a parallel beam of light is incident on a concave lens, we will obtain a diverging spherical wave with $\theta \simeq a/f$ (see Fig. 15.29). Now, Eq. (15.10-8) can be rewritten as

---

*1 Volt/m $\simeq$ 1/3×10^{-4} dynes/esu, where dynes/esu is the CGS unit of the electric field. Thus $n_{2M} = 1/9 \times 10^{-8}\, n_{2C}$ where the subscripts M and C refer to MKS and CGS units respectively.

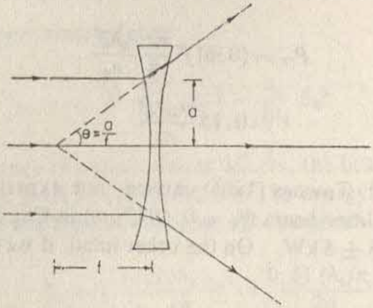


Fig. 15.29 When a plane wave is incident on a diverging lens, the transmitted rays diverge making an angle $\theta \simeq a/f$ with the axis.

$$\theta_d \simeq \frac{a}{fd} \tag{15.10.16}$$

where

$$fd = \frac{a}{\theta_d} \simeq a \frac{2an_0}{0.61\,\lambda_0} \simeq \frac{2\pi}{\lambda_0} n_0 \frac{a^2}{2} = \frac{ka^2}{2} \tag{15.10-17}$$

where

$$k = \frac{2\pi}{\lambda} = \frac{2\pi}{\lambda_0} n_0$$

Thus the phenomenon of diffraction can be approximated by a diverging lens of focal length $ka^2/2$.

Now, we know that, if a converging lens of a certain focal length is placed adjacent to a diverging lens of the same focal length, then a parallel beam incident on such a combination will emerge as a parallel beam, i.e. neither any focusing nor any defocusing will occur. In the present case we have just seen that an incident beam propagates without any divergence or convergence, when

$$E_0^2 = \frac{1}{n_2 n_0}\left[\frac{0.61\,\lambda_0}{2a}\right]^2 \tag{15.10-18}$$

which may be rewritten in the form,

$$f_d = \tfrac{1}{2}\,ka^2 \simeq a\left[\frac{n_0}{2n_2E_0^2}\right]^{\frac{1}{2}} \tag{15.10-19}$$

Thus, the nonlinear medium can be thought of as a converging lens of focal length

$$f_{nl} = a\left[\frac{n_0}{2n_2E_0^2}\right]^{1/2} \tag{15.10 20}$$

When $f_d < f_{nl}$, the diffraction divergence will dominate and when $f_d > f_{nl}$, the focussing action will dominate giving rise to a converging wave in the nonlinear medium.

15.11 Spatial Frequency Filtering

If $g(x, y)$ represents the field distribution on the front focal plane of a corrected lens (i.e., on the plane P_1 in Fig. 15.30), then on the back focal plane P_2 of the lens one obtains the Fourier transform of $f(x, y)$; the z-axis represents the optical axis of the lens. Thus, if $G(x, y)$ represents the field distribution on the back focal plane P_2, then it is related to $g(x, y)$ through the following relation

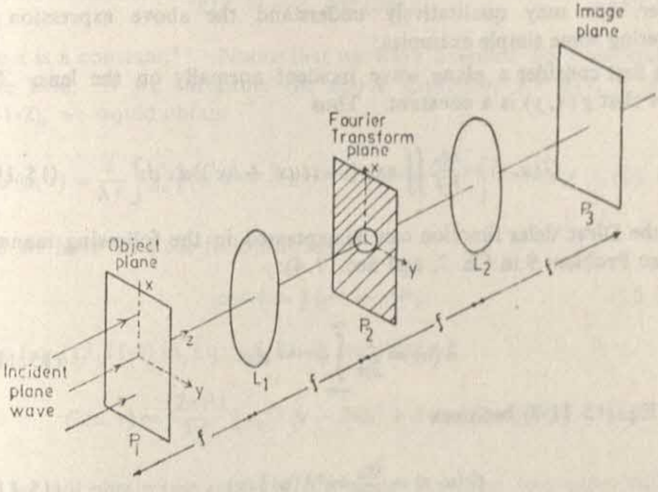


Fig. 15.30 The plane P_2 is the Fourier transform plane where the spatial frequency components of the object (placed in the plane P_1) are displayed. In the above figure, a small hole is placed on the axis (in the plane P_3) which filters out the high frequency components.

$$G(x, y) = \frac{i}{\lambda f} \iint g(x', y') \exp\left[-2\pi i\left(\frac{xx'}{\lambda f} + \frac{yy'}{\lambda f}\right)\right] dx'dy' \quad (15.11\text{-}1)$$

where λ represents the wavelength of the light and f is the focal length of the lens. This remarkable property of a corrected lens forms the basis of the subject of spatial frequency filtering which finds numerous applications in many diverse areas. Equation (15.11-1) is usually written in the form

$$G(u, v) = \frac{i}{\lambda f} \iint g(x', y') \exp\left[-i(ux' + vy')\right] dx'dy' \quad (15.11\text{-}2)$$

where

$$u \equiv \frac{2\pi x}{\lambda f}$$

and

$$v \equiv \frac{2\pi y}{\lambda f}$$

If we compare Eq. (15.11-2) with Eq. (7.4-12), we find that $F(u, v)$ is the Fourier transform of $f(x, y)$ and as such u and v are said to represent the spatial frequencies along the x and y directions respectively. The derivation of Eq. (15.11-2) is fairly involved (see, e.g., Goodman, 1968); however, one may qualitatively understand the above expression by considering some simple examples:

We first consider a plane wave incident normally on the lens. This implies that $g(x, y)$ is a constant. Thus

$$G(u, v) = \frac{ig_0}{\lambda f} \iint \exp\left[-i(ux' + vy')\right] dx' dy' \qquad (15.11\text{-}3)$$

Now, the Dirac delta function can be expressed in the following manner* (see also Problem 9 in Ch. 7, and Sec. 7.4):

$$\delta(u) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-iux'} dx' \qquad (15.11\text{-}4)$$

Thus, Eq. (15.11-3) becomes

$$G(u, v) = \frac{ig_0}{\lambda f} 4\pi^2 \delta(u) \delta(v) \qquad (15.11\text{-}5)$$

where $\delta(u)$ and $\delta(v)$ represent the Dirac delta functions. Since $\delta(u) = 0$ for $u \neq 0$, one can infer from Eq. (15.11-5) that the intensity is zero at all points excepting at the point $x = 0$, $y = 0$. This is to be expected because

---

*For large values of L, the integral

$$I(u, L) = \frac{1}{2\pi} \int_{-L}^{L} e^{-iux'} dx' = \frac{1}{\pi} \frac{\sin(uL)}{u}$$

is a very sharply peaked function around $u = 0$ and is such that

$$\int_{-\infty}^{\infty} I(u, L)\, du = 1 \qquad \text{(independent of } L\text{)}.$$

The Dirac delta function is defined through the following equations:

$$\delta(x - a) = 0 \quad \text{for} \quad x \neq a$$

and

$$\int_{a-\alpha}^{a+\alpha} f(x)\, \delta(x - a)\, dx = f(a)$$

where α is an arbitrary positive quantity and $f(x)$ is a well behaved function. It can be easely seen that as $L \to \infty$, the function $I(u, L)$ behaves as a delta function. Thus

$$\delta(u) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-iux} dx$$

a plane wave gets focussed to a point by a corrected lens.*

Another interesting example is a one-dimensional cosinusoidal field distribution in the object plane, i.e.,

$$g(x, y) = g_0 \cos (2\pi\alpha x) \tag{15.11-6}$$

where α is a constant.** Notice that we have assumed no y dependence of the field. If we substitute the above expression for $g(x, y)$ in Eq. (15.11-2), we would obtain

$$G(u, v) = \frac{i}{\lambda f} \int g_0 \tfrac{1}{2} [e^{i2\pi\alpha x'} + e^{-i2\pi\alpha x'}] \cdot e^{-iux'}\, dx' \int e^{-ivy'} dy' \tag{15.11-7}$$

where we have used the following identity:

$$\cos \theta = \tfrac{1}{2} [e^{i\theta} + e^{-i\theta}] \tag{15.11-8}$$

Using Eq. (15.11-4) in Eq. (15.11-7), we obtain

$$G(u, v) = \frac{(2\pi)^2 i}{\lambda f} \tfrac{1}{2} g_0 [\delta (u - 2\pi\alpha) + \delta (u + 2\pi\alpha)] \delta (v) \tag{15.11-9}$$

Thus, we will obtain two spots in the plane P_2; these two spots will be lying on the x-axis at $u = \pm 2\pi\alpha$ (i.e., at $x = \pm \lambda f \alpha$). Physically this can be understood from the following consideration: When a plane wave is incident normally on the plane P_1, (see Fig. 15.30), the time dependence is of the form $\cos \omega t$. If in the plane P_1, we have an object whose transmittance is proportional to $\cos (2\pi\alpha x)$, then the field to the right of the plane P_1 would be proportional to

$$\cos \omega t \cos 2\pi\alpha x = \tfrac{1}{2} [\cos (\omega t + 2\pi\alpha x) + \cos (\omega t - 2\pi\alpha x)] \tag{15.11-10}$$

We know that for a plane wave having no phase variation in the y-direction, the field is of the form (see Example 8.6):

$$\cos (\omega t + k_x x + k_z z) \tag{15.11-11}$$

where $k_x = k \sin \theta$, $k_z = k \cos \theta$, $k = 2\pi/\lambda$ and θ is the angle that the propagation vector **k** makes with the z-axis. At $z = 0$, the field becomes

$$\cos (\omega t + k_x x) \tag{15.11-12}$$

---

*We are assuming here a very large dimension of the aperture of the lens; as such, the limits of integrations in Eq. (15.11-2), are assumed to be from $-\infty$ to $+\infty$. This is a reasonably good approximation for most practical cases.

**On the plane P_1 (see Fig. 15.30) if we place the negative of the photograph shown in Plate 6 with the y-axis along the length of a fringe and assume a plane wave to be incident normally on the film, then the field distribution would be proportional to $\cos^2 \alpha x$ which is equal to $\frac{1}{2} (1 + \cos 2\alpha x)$.

Comparing Eqs. (15.11-10) and Eq. (15.11-12), we find that the two terms on the RHS of Eq. (15.11-10) represent two plane waves propagating along directions making angles $+\theta$ and $-\theta$ with the z-axis, where

$$\sin\theta = \frac{k_x}{k} = \frac{2\pi\alpha}{2\pi/\lambda} = \alpha\lambda \tag{15.11-13}$$

These plane waves will obviously focus to two points at $x = +\lambda f\alpha$ and $x = -\lambda f\alpha$ on the x-axis in the plane P_2. Since α represents the spatial frequency associated with the object, one essentially obtains the spatial frequency spectrum of the object on the back focal plane.

We are familiar with the fact that a general time varying signal can be expressed as a superposition of pure sinusoidal signals (see Eq. 7.4-13); in a similar manner, the field variation across an arbitrary object (placed in the plane P_1), can be expressed as a superposition of sinusoidal variations and one would get the corresponding frequency components on the plane P_2. For this reason, the plane P_2 is often termed as the Fourier transform plane.

As another example, if the amplitude variation of the object is of the form

$$g(x, y) = A\cos 2\pi\alpha x + B\cos 2\pi\beta x \tag{15.11-14}$$

then one would obtain four spots in the plane P_1 (all lying on the x-axis); these spots will appear at $x = \pm\lambda f\alpha, \pm\lambda f\beta$. Since the Fourier transform of the Fourier transform is the original function itself* [see Eq. (7.4-11)], if we place the plane P_2 on the front focal plane of another lens L_2, then on its back focal plane (i.e., in the plane P_3 in Fig. 15.30), we will obtain the amplitude distribution associated with the object. If we now put stops at the points $(x = +\lambda f\alpha, y = 0)$ and $(x = -\lambda f\alpha, y = 0)$ in the plane P_2, then the field distribution in the plane P_3 would be proportional to $\cos 2\pi\beta x$. Thus we have been able to filter out the spatial frequency α. This is the basic principle behind spatial frequency filtering.

For an arbitrary object if we put a small stop on the axis, then it will filter out the low frequency components; on the other hand, if we put a small hole on the axis (see Fig. 15.30), we filter out the high frequency components.

As a simple application, we consider a half-tone photograph (like that in a newspaper) which consists of a large number of spots of varying shades which produce the image pattern. Since the spots are spaced closely, it represents a high frequency noise and the overall image has much smaller frequencies associated with it. Thus if we put a transparency similar to that shown in Plate 15a, and allow only the low frequency components to pass through (as shown in Fig. 15.30), we will obtain, in the plane P_3 an image which

---

*There will however be an inversion; i.e., $f(x, y)$ will become $f(-x, -y)$ in the plane P_3. This can also be seen by simple ray tracing.

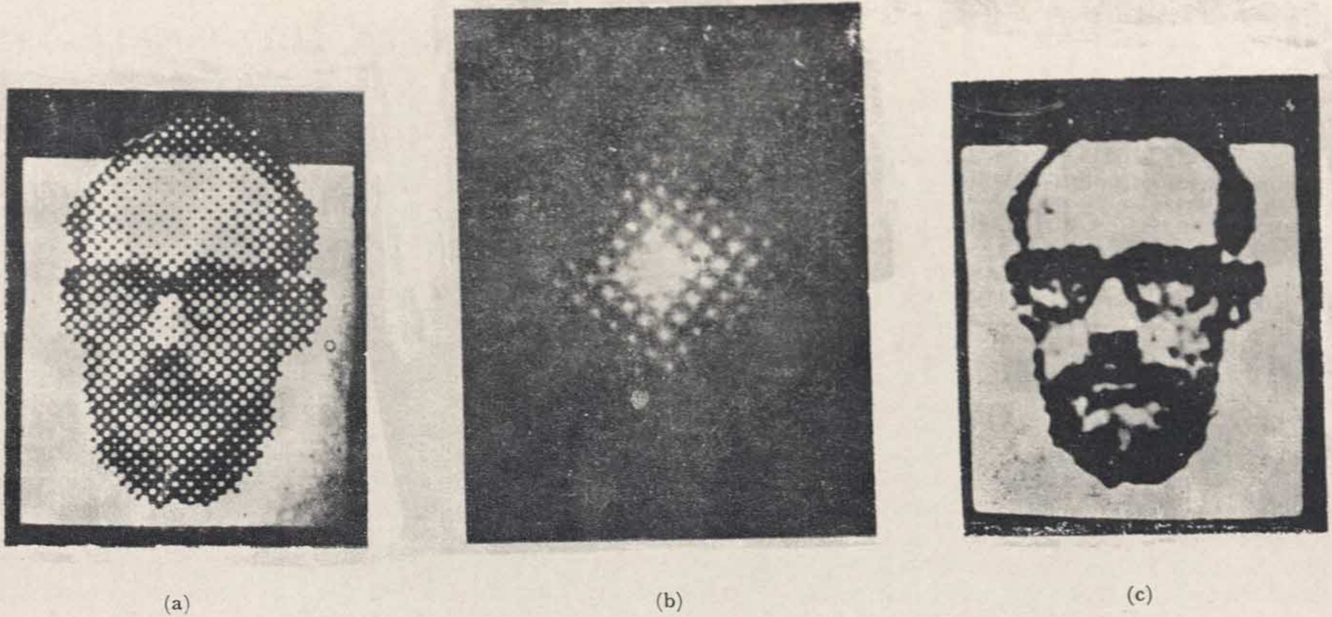

(a) (b) (c)

Plate 15 (a) Shows a photograph consisting of regularly spaced black and white squares of varying sizes. When a pinhole is placed in the Fourier transform plane to block the high frequency components an image of the form shown in (c) is obtained; the frequency spectrum is shown in (b). Notice that in (c) shades of gray appear as well as details such as the missing part of the eye glass frame. (After R.A. Phillips, "Spatial Filtering Experiments for Undergraduate Laboratories," *American Journal* [illegible] does not contain the unwanted high frequency noise (see Plate 15c).

The subject of spatial frequency filtering finds applications in many other areas like contrast improvement, character recognition etc.

PROBLEMS

1. A plane wave ($\lambda = 5000$ Å) falls normally on a long narrow slit of width 0.5 mm. Calculate the angles of diffraction corresponding to the first three minima. Repeat the calculations corresponding to a slit width of 0.1 mm. Interpret physically the change in the diffraction pattern.
2. A convex lens of focal length 20 cm is placed after a slit of width 0.6 mm. If a plane wave of wavelength 6000 Å falls normally on the slit, calculate the separation between the second minima on either side of the central maxima. (Ans. $\simeq 0.08$ cm).
3. In the above problem calculate the ratio of the intensity of the principal maximum to the first maximum on either side of the principal maximum.
4. Consider a laser beam of circular cross-section of diameter 3 cm and of wavelength 5×10^{-5} cm. Calculate the order of the beam diameter after it has traversed a distance of 3 km. (Ans. ~ 5 cm. This shows the extremely high directionality of laser beams).
5. A circular aperture of radius 0.01 cm is placed in front of a convex lens of focal length 25 cm and illuminated by a parallel beam of light of wavelength 5×10^{-5} cm. Calculate the radii of the first three dark rings.
6. Consider a plane wave incident on a convex lens (of diameter 5 cm) and of focal length 10 cm. If the wavelength of the incident light is 6000 Å, calculate the radius of the first dark ring on the focal plane of the lens. Repeat the calculations for a lens of same focal length but diameter 15 cm. Interpret the results physically.
7. Consider a set of two slits each of width $b = 5 \times 10^{-2}$ cm and separated by a distance $d = 0.1$ cm, illuminated by a monochromatic light of wavelength 6.328×10^{-5} cm. If a convex lens of focal length 10 cm is placed beyond the double slit arrangement, calculate the positions of the maxima inside the first diffraction minimum.
8. Show that when $b = d$, the resulting diffraction pattern corresponds to a slit of width $2b$.
9. Calculate the maximum value of the grating element d so that the first order and second order spectra do not overlap when the grating is used for studying a light beam containing wavelength components from 4000 Å to 7000 Å.
10. Consider a diffraction grating of width 5 cm with slits of width 0.0001 cm separated by a distance of 0.0002 cm. What is the corresponding grating element? How many orders would be observable at $\lambda = 5.5 \times 10^{-5}$ cm? Calculate the width of the principal maximum. Would there be any missing orders?
11. For the diffraction grating of Problem 10, calculate the dispersion in the different orders. What will be the resolving power in each order?
12. A grating (with 15,000 lines per inch) is illuminated by white light. Assuming that white light consists of wavelengths lying between 4000 and 7000 Å, calculate the angular widths of the first and the second order spectra. (Hint: You should not use Eq. (15.7-2): why?).
13. A grating (with 15,000 lines per inch) is illuminated by sodium light. The grating spectrum is observed on the focal plane of a convex lens of focal length 10 cm. Calculate the separation between the D_1 and D_2 lines of sodium. (The wave-

length of the D_1 and D_2 lines are 5890 and 5896 Å respectively). (Hint: You will have to use Eq. (15.7–2)).

14. Calculate the resolving power in the second order spectrum of a 1 inch grating having 15,000 lines.
15. Consider a wire grating of width 1 cm having 1000 wires. Calculate the angular width of the second order principal maxima and compare the value with the one corresponding to a grating having 5000 lines in 1 cm. Assume $\lambda = 5 \times 10^{-5}$ cm.
16. In the minimum deviation position of a diffraction grating the first order spectrum corresponds to an angular deviation of 30°. If $\lambda = 6 \times 10^{-5}$ cm, calculate the grating element.
17. Calculate the diameter of a telescope lens if a resolution of 0.1 seconds of arc is required at $\lambda = 6 \times 10^{-5}$ cm.
18. Assuming that the resolving power of the eye is determined by diffraction effects only, calculate the maximum distance at which two objects separated by a distance of 2 m can be resolved by the eye. (Assume pupil diameter to be 2 mm and $\lambda = 6000$ Å).
19. Copper is an fcc structure with lattice constant 3.615 Å. An X-ray powder photograph of copper is taken. The X-ray beam consists of wavelengths 1.540 Å and 1.544 Å. Show that diffraction maxima will be observed at $\theta = (21.64°, 21.70°), (25.21°, 25.28°), (37.05°, 37.16°), (44.94°, 45.09°), (47.55°, 47.71°), (58.43°, 58.67°), (68.20°, 68.58°), (72.29°, 72.76°)$.
20. Tungsten is a bcc structure with lattice constant 3.1648 Å. Show that in the powder photograph of tungsten (corresponding to an X-ray wavelength of 1.542 Å) one would observe diffraction maxima at $\theta = 20.15°$, 29.17°, 36.64°, 43.56°, 50.39°, 57.55°, 65.74° and 77.03°.
21. (a) In the simple cubic structure if we alternately place Na and Cl atoms we would obtain the NaCl structure. Show that the Na atoms (and the Cl atoms) independently form fcc structures. The lattice constant associated with each fcc structure is 5.6402 Å. Corresponding to the X-ray wavelength 1.542 Å, show that diffraction maxima will be observed at $\theta = 13.69°$, 15.86°, 22.75°, 26.95°, 28.27°, 33.15°, 36.57°, 37.69°, 42.05°, 45.26°, 50.66°, 53.98°, 55.10°, 59.84°, 63.69°, 65.06°, 71.27°, 77.45° and 80.66°.

    (b) Show that if we treat NaCl as a simple cubic structure with lattice parameter 2.82 Å then the maxima at $\theta = 13.69°$, 26.95°, 36.57°, 45.26°, 53.98°, 63.69° and 77.45° will not be observed. Indeed in the X-ray diffraction pattern of NaCl, the maxima corresponding to these angles will be very weak.
22. Calculate the Fraunhofer diffraction pattern produced by a double slit arrangement with slits of widths b and $3b$, with their centres separated by a distance $6b$.
23. Consider the propagation of a 1 kW laser beam ($\lambda = 6943$ Å, beam diameter $\simeq 1$ cm) in CS_2. Calculate f_d and f_{nl} and discuss the defocussing (or focussing) of the beam. Repeat the calculations corresponding to a 1000 kW beam and discuss any qualitative differences that exist between the two cases. The data for n_0 and n_2 are given in Sec. 15.10.
24. The values of n_0 and n_2 for benzene are 1.5 and 0.6×10^{-10} C.G.S. units respectively. Obtain an approximate expression for the critical powers.

REFERENCES AND SUGGESTED READING

See at the end of Chapter 16.

16 Fresnel Diffraction

> *Woe to the man who seeks to shed a brilliant light in a place where people want to keep in darkness and shadow.*
>
> BENEDETTO CROCE

16.1 Introduction

In the previous chapter we had mentioned that the phenomenon of diffraction can be broadly classified under two categories: under the first category comes the Fresnel class of diffraction in which either the source or the screen (or both) are at a finite distance from the diffracting aperture. In the second category comes the Fraunhofer class of diffraction (discussed in the previous chapter) in which the wave incident on the aperture is a plane wave and the diffraction pattern is observed on the focal plane of a convex lens, so that the screen is effectively at an infinite distance from the aperture. In the present chapter we will discuss the Fresnel class of diffraction and also study the transition to the Fraunhofer region. The underlying principle in the entire analysis is the Huygens-Fresnel principle according to which each point on a wave-front is a source of secondary disturbance and the secondary wavelets emanating from different points mutually interfere. In order to appreciate the implications of this principle we consider the following example.

Let us consider the incidence of a plane wave on a circular hole of diameter D as shown in Fig. 16.1. In Sec. 15.3 we had shown that the beam will undergo diffraction divergence and the angular spreading $\Delta\theta$ will be $\sim\lambda/D$. Thus, when $D>>>\lambda$ the intensity at a point P (which is deep inside the geometrical shadow) will be negligible; on the other hand, if $D \lesssim \lambda$ there will be almost uniform spreading out of the beam resulting in an (almost) uniform illumination of the screen. This phenomenon is a manifestation of the fact that when $D >>> \lambda$, the secondary wavelets emanating from different points on the circular aperture, so beautifully interfere to produce (almost) zero intensity in the geometrical shadow and a large intensity in the region AB (see Fig. 16.1). However, if $D \lesssim \lambda$ then the aperture almost acts as a point source resulting in a uniform illumination of the screen (see Plate 4).

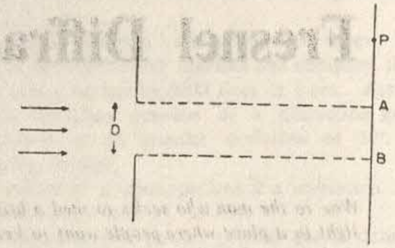


Fig. 16.1 Diffraction of plane waves by a circular aperture.

We will first introduce the concept of Fresnel half-period zones to have a qualitative understanding of the Fresnel diffraction pattern; this will be followed by a more rigorous analysis of the Fresnel class of diffraction and its transition to the Fraunhofer region.

16.2 Fresnel Half-period Zones

Let us consider a plane wavefront WW' propagating in the z-direction as shown in Fig. 16.2. In order to determine the field at an arbitrary

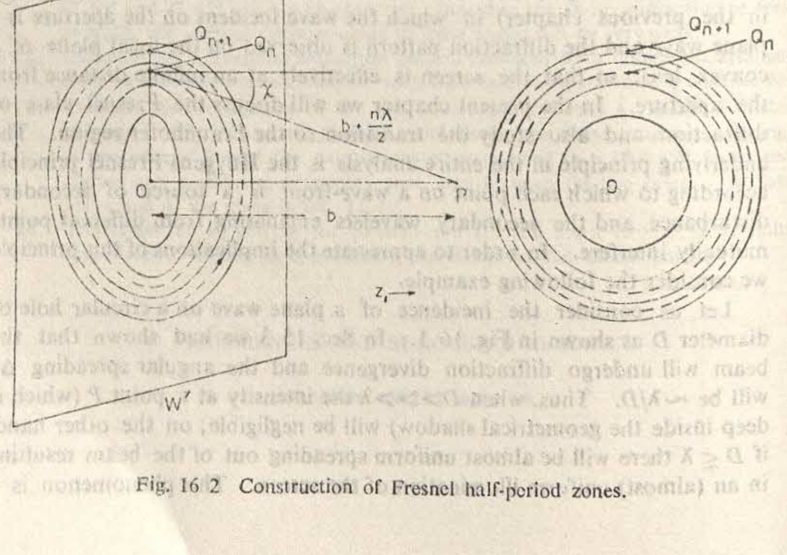


Fig. 16 2 Construction of Fresnel half-period zones.

point P due to the disturbances reaching from different portions of the wavefront, we make the following construction: from the point P we drop a perpendicular PO on the wavefront. If $PO = b$, then with point P as centre we draw spheres of radii $b, b+\frac{\lambda}{2}, b+\frac{2\lambda}{2}, b+\frac{3\lambda}{2}, \ldots$; these spheres will intersect WW' in circles as shown in Fig. 16.2. The radius of the nth circle will obviously be given by

$$r_n = \left[\left(b + n\frac{\lambda}{2}\right)^2 - b^2\right]^{1/2} = \sqrt{nb\lambda}\left[1+\frac{n\lambda}{4b}\right]^{1/2} \qquad (16.2\text{-}1)$$

or

$$r_n \simeq \sqrt{nb\lambda} \qquad (16.2\text{-}2)$$

where we have neglected the term $\frac{n^2\lambda^2}{4}$ in comparison to $nb\lambda$; this will indeed be justified for practical systems using visible light where $b >>> \lambda$. The area of the annular region between two consecutive circles will be

$$\pi r_{n+1}^2 - \pi r_n^2 \simeq \pi[(n+1)b\lambda - nb\lambda]$$
$$\simeq \pi b\lambda$$

Thus, the areas of all the annular regions are approximately equal. The annular region between the $(n-1)^{\text{th}}$ and the n^{th} circle is known as the n^{th} zone. Now, the resultant disturbance produced by the n^{th} zone will be π out-of-phase with the disturbance produced by the $(n-1)^{\text{th}}$ [or the $(n+1)^{\text{th}}$] zone. This can be seen from the fact that for an arbitrary point Q_n in the n^{th} half-period zone there will be a corresponding point Q_{n+1} in the $(n+1)^{\text{th}}$ zone such that $Q_{n+1}P - Q_nP = \frac{\lambda}{2}$, which corresponds to a phase difference of π. Since the areas of the zones are approximately equal, one can have a one-to-one correspondence between points in various zones. Thus, the resultant amplitude at the point P can be written as

$$u(P) = u_1 - u_2 + u_3 - u_4 + \ldots + (-1)^{m+1} u_m \qquad (16.2\text{-}3)$$

where u_n represents the net amplitude produced by the secondary wavelets emanating from the n^{th} zone; the alternate negative and positive signs represent the fact that the resultant disturbances produced by two consecutive zones are π out of-phase with respect to each other. It may be pointed out that the amplitude produced by a particular zone is proportional to the area of the zone and inversely proportional to the distance of the zone from the point P; further, it also depends on an obliquity factor* which is proportional to $\frac{1}{2}(1+\cos\chi)$ where χ is the angle that the normal to the zone makes with the line QP. Thus we may write

---

*This obliquity factor automatically comes from rigorous diffraction theory; see, for example, Born and Wolf (1975).

$$u_n = \text{constant}\ \frac{A_n}{Q_n P}\ \frac{(1+\cos\chi)}{2} \tag{16.2-4}$$

where A_n represents the area of the n^{th} zone. It can be shown that if we use Eq. (16.2-1) for r_n, the areas of the zones increase with n; however, this slight increase in the area is exactly compensated by the increased distance of the zone from the point P. Inspite of this, the amplitudes $u_1, u_2, u_3, \ldots$ decrease monotonically because of increased obliquity. Thus we may write

$$u_1 > u_2 > u_3 \ldots \tag{16.2-5}$$

The series expressed by Eq. (16.2-3) can be approximately summed due to a method by Schuster. We rewrite Eq. (16.2-3) as

$$u(P) = \frac{u_1}{2} + \left(\frac{u_1}{2} - u_2 + \frac{u_3}{2}\right) + \left(\frac{u_3}{2} - u_4 + \frac{u_5}{2}\right) + \ldots \tag{16.2-6}$$

where the last term would be either $\frac{1}{2}u_m$ or $(\frac{1}{2}u_{m-1} - u_m)$ according to m being odd or even. If the obliquity factor is such that

$$u_n > \tfrac{1}{2}(u_{n-1} + u_{n+1}) \tag{16.2-7}$$

then the quantities inside the brackets in Eq. (16.2-6) will be negative; consequently

$$\left.\begin{aligned} &u(P) < \tfrac{1}{2}u_1 + \tfrac{1}{2}u_m \,;\ (m \text{ odd}) \\ &\text{and} \\ &u(P) < \tfrac{1}{2}u_1 + \tfrac{1}{2}u_{m-1} - u_m \approx \frac{u_1}{2} - \frac{u_m}{2}\,;\ (m \text{ even}) \end{aligned}\right\} \tag{16.2-8}$$

where we have assumed that the amplitude of the fields produced by consecutive zones differ only slightly. In order to obtain the upper limits we rewrite Eq. (16.2-3) in the form

$$u(P) = u_1 - \frac{u_2}{2} - \left(\frac{u_2}{2} - u_3 + \frac{u_4}{2}\right) - \left(\frac{u_4}{2} - u_5 + \frac{u_6}{2}\right)\ldots \tag{16.2-9}$$

where the last term would now be $-\frac{1}{2}u_{m-1} + u_m$ when m is odd and $-\frac{1}{2}u_m$ when m is even. Since the quantities inside the brackets are negative we obtain

$$\left.\begin{aligned} &u(P) > u_1 - \frac{u_2}{2} - \frac{u_{m-1}}{2} + u_m \approx \frac{u_1}{2} + \frac{u_m}{2} \quad \text{when } m \text{ is odd} \\ &\text{and} \\ &u(P) > u_1 - \frac{u_1}{2} - \frac{u_m}{2} \approx \frac{u_1}{2} - \frac{u_m}{2} \quad \text{when } m \text{ is even} \end{aligned}\right\} \tag{16.2-10}$$

Using Eqs. (16.2-8) and (16.2-10) we may approximately write

$$\left.\begin{aligned} u(P) &= \frac{u_1}{2} + \frac{u_m} {2} \quad \text{when } m \text{ is odd} \\ \text{and} \qquad\qquad & \\ u(P) &= \frac{u_1}{2} - \frac{u_m}{2} \quad \text{when } m \text{ is even} \end{aligned}\right\} \tag{16.2-11}$$

If we can neglect* u_m in comparison to u_1 then the above equation gives the remarkable result that

$$u(P) \approx \frac{u_1}{2} \tag{16.2-12}$$

implying that the resultant amplitude produced by the entire wavefront is only one half of the amplitude produced by the first half period zone.

We may use the above analysis to study the diffraction of a plane wave by a circular aperture. Let the point P be at a distance b from the circular aperture (see Fig. 16.3). We assume that the radius of the circular

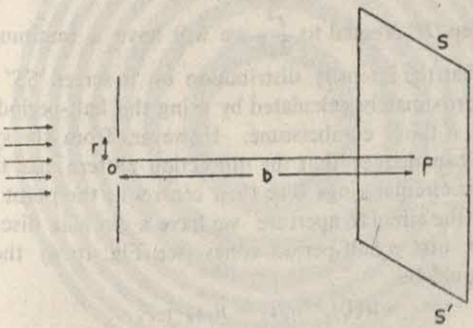


Fig. 16.3 Diffraction of a plane wave by a circular aperture.

aperture, r, can be increased from zero onwards. As r increases the intensity at the point P would also increase till the circular aperture contains the first half-period zone; this would happen when $r = \sqrt{b\lambda}$. The resultant amplitude at the point P would be u_1 which is twice the value of the amplitude for the unobstructed wavefront [see Eq. 16.2-12)]. The intensity would therefore be $4I_0$, where I_0 represents the intensity at the point P due to the unobstructed wavefront. If we further increase r then $u(P)$ would start decreasing and when the circular aperture contains the first two half period zones (which would happen when $r=\sqrt{2b\lambda}$) the resulant amplitude ($=u_1-u_2$) would be almost zero. Thus, by increasing the hole diameter, the intensity at the point P decreases almost to zero. This interesting result is once again due to the validity of the Huygens-Fresnel principle and hence would be valid for sound waves also. We may generalize the above result by noting that if $r=\sqrt{(2n+1)\,b\lambda}$ (i.e. if the aperture contains an odd number of half-period zones) then the intensity will be maximum and if $r=\sqrt{2nb\lambda}$ the intensity will be minimum. In order to have a numerical appreciation, we note that for $b=50$ cm and $\lambda=5\times10^{-5}$ cm, the radii of the first, second and third zones would be 0.500 mm, 0.707 mm and 0.866 mm respectively. As a corollary of the above analysis we can consider a circular aperture of a fixed radius and study the intensity variation along the axis. Whenever the distance OP is equal to $\frac{r^2}{(2n+1)\lambda}$, the point P (see Fig. 16.3) would correspond to a maximum and when OP is equal to $\frac{r^2}{2n\lambda}$ we will have a minimum. It may be pointed out that the intensity distribution on a screen SS' at off-axis points can be approximately calculated by using the half-period zones but, such a calculation is fairly cumbersome. However, from the symmetry of the problem, one can deduce that the diffraction pattern has to be in the form of concentric circular rings with their centres at the point P.

---

*If one assumes a form of the obliquity factor as given by Eq. (16.2-4) then it decreases from 1 to $\frac{1}{2}$ as m increases from 1 to ∞, this implies that $|u_m|$ can never be smaller than $\frac{u_1}{2}$. However, when m is large, a slight shift of the point P on the axis will change the amplitude from $\frac{u_1}{2} + \frac{u_m}{2}$ to $\frac{u_1}{2} - \frac{u_m}{2}$; the changes will occur with such great rapidity that one can only observe the average which will be $\frac{u_1}{2}$.

If instead of the circular aperture we have a circular disc and if the disc obstructs the first p half-period zones (see Fig. 16.4) then the field at the point P would be

$$u(P)=u_{p+1}-u_{p+2}+\dots$$

$$\simeq\frac{u_{p+1}}{2}\qquad(16.2\text{-}13)$$

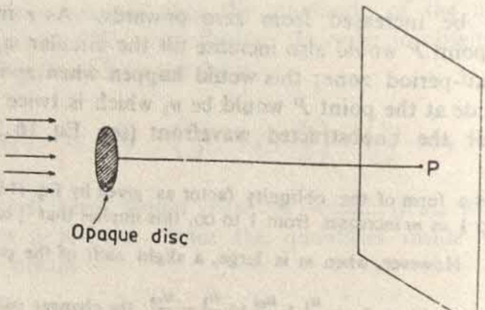


Fig. 16.4 When a plane wave is incident on an opaque disc then a bright spot will be formed on an axial point. This spot is known as the Poisson spot.

Thus, we should always obtain a bright spot on the axis, which has indeed been confirmed by experiments. The more rigorous theory also predicts the same result (see Sec. 16.8).

16.3 The Zone-plate

A beautiful application of the concept of Fresnel half-period zones lies in the construction of the zone-plate which consists of a large number of concentric circles whose radii are proportional to the square root of natural numbers and the alternate annular regions of which are blackened (see Fig. 16.5). Let the radii of the circles be $\sqrt{1K}$, $\sqrt{2K}$, $\sqrt{3K}$, $\sqrt{4K}$,. . . .,

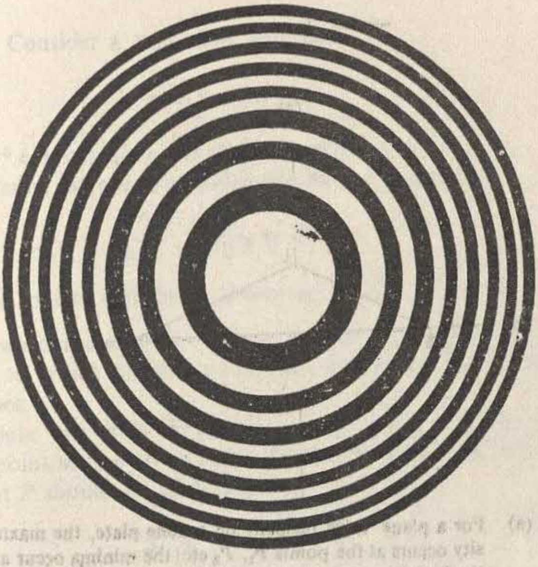


Fig. 16.5 The zone plate.

where K is a constant. We consider a point P_1 which is at a distance K^2/λ from the zone plate; for this point the blackened rings correspond to the 2nd, 4th, 6th. . . half-period zones. Thus, the even zones are obstructed and the resultant amplitude at P_1 [see Fig. 16.6(a)] will be

$$u_1 + u_3 + u_5 + \cdots \tag{16.3-1}$$

producing an intense maximum. For the point P_3 (which is at a distance $K^2/3\lambda$) the first blackened ring contains the 4th, 5th, 6th zones, the second blackened ring contains the 10th, 11th and 12th zones, etc.; thus the resultant amplitude would be

$$(u_1 - u_2 + u_3) + (u_7 - u_8 + u_9) + \cdots \tag{16.3-2}$$

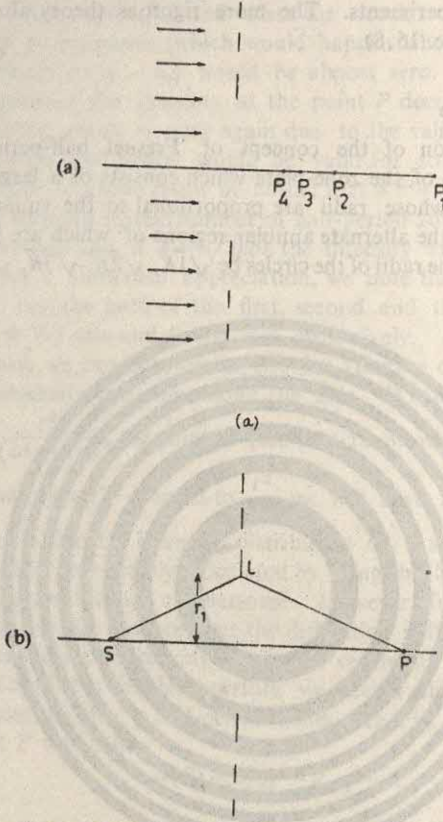


Fig. 16.6 (a) For a plane wave incident on a zone plate, the maxiumm intensity occurs at the points P_1, P_3 etc; the minima occur at $P_2, P_4, \ldots$ (b) Imaging of a point object by a zone-plate.

which would again correspond to a maximum, but it would not be as intense as the point P_1. Between the points P_1 and P_3 there will be a point P_2 (at a distance $K^2/2\lambda$) where the resultant amplitude would be

$$(u_1 - u_2) + (u_5 - u_6) + \ldots \qquad (16.3\text{-}3)$$

implying that corresponding to P_2 the first blackened ring contains the 3rd and 4th half-period zones, etc. Obviously, the point P_2 will correspond to a minimum. Thus, if a plane wave is incident normally on a zone-plate, then the corresponding focal points are at distances

$$\frac{K^2}{\lambda}, \frac{K^2}{3\lambda}, \frac{K^2}{5\lambda}, \ldots \qquad (16.3\text{-}4)$$

from the zone-plate.

Example 16.1 Assume a plane wave ($\lambda = 5\times10^{-5}$ cm) to be incident on a circular aperture of radius 0.5 mm. Calculate the positions of the brightest and darkest points on the axis.

Solution. For the brightest point the aperture should contain only the first zone and thus we must have (see Fig. 16.3)

$$(0.05)^2 = OP(5\times10^{-5})$$

Thus

$$OP = 50 \text{ cm}$$

Similarly the darkest point would be at a distance

$$\frac{(0.05)^2}{2\times5\times10^{-5}} = 25 \text{ cm}$$

Example 16.2 Consider a zone plate with radii

$$r_n = 0.1\sqrt{n} \text{ cm}$$

For $\lambda = 5\times10^{-5}$ cm, calculate the positions of the foci.

Solution. The most intense focal point will be at a distance

$$\frac{r_1^2}{\lambda} = \frac{0.01}{5\times10^{-5}} = 200 \text{ cm}$$

The other focal points will be at distances of $\frac{200}{3}$, $\frac{200}{5}$, $\frac{200}{7}$ cm, etc. Between any two consecutive foci there will be dark points on the axis corresponding to which the first circle will contain an even number of half-period zones.

The zone-plate can also be used for imaging points on the axis, e.g. if we have a point source at S then a bright image will be formed at P, where the point P should be such that

$$SL + LP - SP = \frac{\lambda}{2} \tag{16.3-5}$$

the point L being on the periphery of the first circle of the zone plate [see Fig. 16.6 (b)]. If the radius of the first circle is r_1 then

$$\begin{aligned} SL + LP - SP &= \sqrt{a^2 + r_1^2} + \sqrt{b^2 + r_1^2} - (a+b) \\ &\simeq a\left[1 + \frac{r_1^2}{2b^2}\right] + b\left[1 + \frac{r_1^2}{2a^2}\right] - (a+b) \\ &\simeq \frac{r_1^2}{2}\left(\frac{1}{a} + \frac{1}{b}\right) \end{aligned} \tag{16.3-6}$$

Thus Eq. (16.3-5) becomes

$$\frac{1}{a} + \frac{1}{b} = \frac{1}{f} \tag{16.3-7}$$

where $f = r_1^2/\lambda$ represents the focal length. Eq. (16.3-7) resembles the lens law.

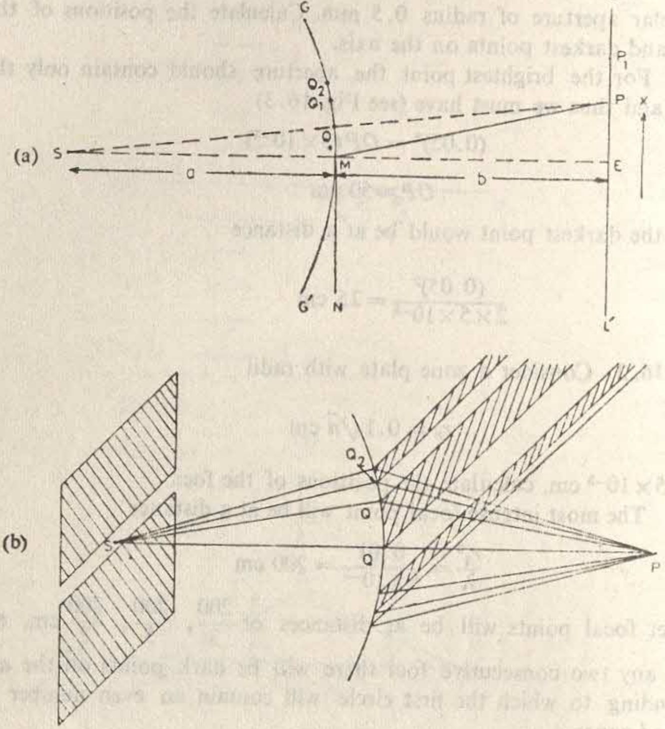


Fig. 16.7 (a) Diffraction at a straight edge. (b) Half-period strips of a cylindrical wavefront.

16.4 Diffraction at a Straight Edge

Let us consider a straight edge *MN* placed perpendicular to the plane of the paper and parallel to a long narrow slit *S* (see Fig. 16.7). We wish to calculate the intensity variation on the screen *LL'*. From the geometry of the arrangement it is obvious that on the screen there will be no intensity variation along the direction parallel to the length of the slit. Thus, the fringes (wherever they occur) will be straight lines parallel to the edge.

The wavefront emanating from the slit is cylindrical and in order to find the amplitude at an arbitrary point *P* (on the screen), we draw half-period strips in the following manner: Let *GQMG'* represent a section of the wavefront, the point *Q* lies on the line joining *S* and *P*. The points Q_1, Q_2 on the wavefront are such that

$$\left.\begin{aligned} SQ_1 + Q_1P - SQP &= \frac{\lambda}{2} \\ SQ_2 + Q_2P - SQP &= \frac{2\lambda}{2}, \text{ etc.} \end{aligned}\right\} \qquad (16.4\text{-}1)$$

The half-period strips will be on the surface of the cylindrical wavefront as shown in Fig. 16.7 (b). However, unlike the Fresnel half-period zones, the areas of the half-period strips will not be equal and thus the analysis becomes quite difficult. Even then one can draw the following conclusions:

(i) Corresponding to the edge of the geometrical shadow (which is shown as E), half of the wavefront is obstructed by the edge, hence the amplitude will be given by

$$u(E) = \tfrac{1}{2}u_0 \tag{16.4-2}$$

where u_0 represents the amplitude that would be produced by the unobstructed wavefront (i.e. in the absence of the edge). Thus the intensity will be given by

$$I(E) = \tfrac{1}{4}I_0 \tag{16.4-3}$$

(ii) Let us next assume that the point P satisfies the following relation:

$$SM + MP - SQP = \frac{\lambda}{2} \tag{16.4-4}$$

Thus only the first half-period strip of the lower part of the wavefront contributes and the resultant amplitude would approximately be

$$\frac{u_1}{2} + \frac{u_1}{4} = \frac{3u_1}{4} = \frac{3}{2}\left(\frac{u_1}{2}\right) = \frac{3}{2}\,u_0 \tag{16.4-5}$$

where $\frac{u_1}{2}$ is the amplitude produced by the first half period strip in the lower portion and $\frac{u_1}{4}$ is the resultant amplitude produced by the upper half of the wavefront (see Eq. 16.2-12). The intensity would be $\frac{9}{4}I_0$. For a point P_1 such that

$$SM + MP_1 - SP_1 = \lambda \tag{16.4-6}$$

we will have a minimum and the resultant amplitude will be

$$\left(\frac{u_1}{2} - \frac{u_2}{2}\right) + \frac{u_1}{4} \tag{16.4-7}$$

In general, an arbitrary point P will correspond to maximum intensity if

$$SM + MP - SP = (2n+1)\frac{\lambda}{2}\,;\ n = 0, 1, 2, \ldots \tag{16.4-8}$$

and a minimum if

$$SM + MP - SP = 2n\,\frac{\lambda}{2}\,;\ n = 1, 2, \ldots \tag{16.4-9}$$

Now,

$$MP = [b^2 + x^2]^{1/2} \simeq b\left(1 + \frac{1}{2}\frac{x^2}{b^2}\right) = b + \frac{x^2}{2b}$$

$$SP = [(a+b)^2 + x^2]^{1/2} \simeq (a+b) + \frac{x^2}{2(a+b)}$$

Hence,

$$SM + MP - SP \simeq a + b + \frac{x^2}{2b} - (a+b) - \frac{x^2}{2(a+b)} = \frac{a}{2(a+b)\,b}\,x^2$$

Thus, when

$$x \cong \left[(2n+1)\frac{b(a+b)}{a}\lambda\right]^{1/2};\ n = 0, 1, 2, \ldots \qquad (16.4\text{-}10)$$

we will have a maximum. For example, for $a = b = 25$ cm and $\lambda = 5 \times 10^{-5}$ cm,

$$\left[\frac{b(a+b)}{a}\lambda\right]^{1/2} = 5 \times 10^{-2} \text{ cm.}$$

Thus the first maximum will occur at a distance of 0.05 cm from the edge of the shadow and the second and third maxima will occur at distances of 0.0866 cm and 0.112 cm respectively. The distance between two consecutive maxima will decrease as we go away from the edge of the geometrical shadow.

Similarly, the positions of the minima are given by

$$x = \left[2n\frac{b(a+b)}{a}\lambda\right]^{1/2};\ n = 1, 2, 3, \ldots \qquad (16.4\text{-}11)$$

and for the above parameters they will occur at distance of 0.07 cm, 0.10 cm, etc. By determining the positions of these maxima and minima one can calculate the wavelength. The precise variation of the intensity is difficult to calculate from this analysis; from a more rigorous theory one obtains a curve as shown in Fig. 16.10.

(iii) If we go inside the geometrical shadow the intensity would continuously decrease as shown in Fig. 16.10.

16.5 Diffraction of a Plane Wave by a Straight Edge—A More Rigorous Approach

In the previous section we had qualitatively discussed the Fresnel diffraction pattern produced by a straight edge. In this section we will make a more detailed (and rigorous) analysis of the diffraction of a plane wave by a straight edge (see Fig. 16.8). Once again, there will be no variation of intensity along the *x*-axis and, therefore, without any loss of generality, we may assume the co-ordinates of an arbitrary point *P* (lying on the screen) to be (0, *y*), where the origin has been assumed to be on the edge

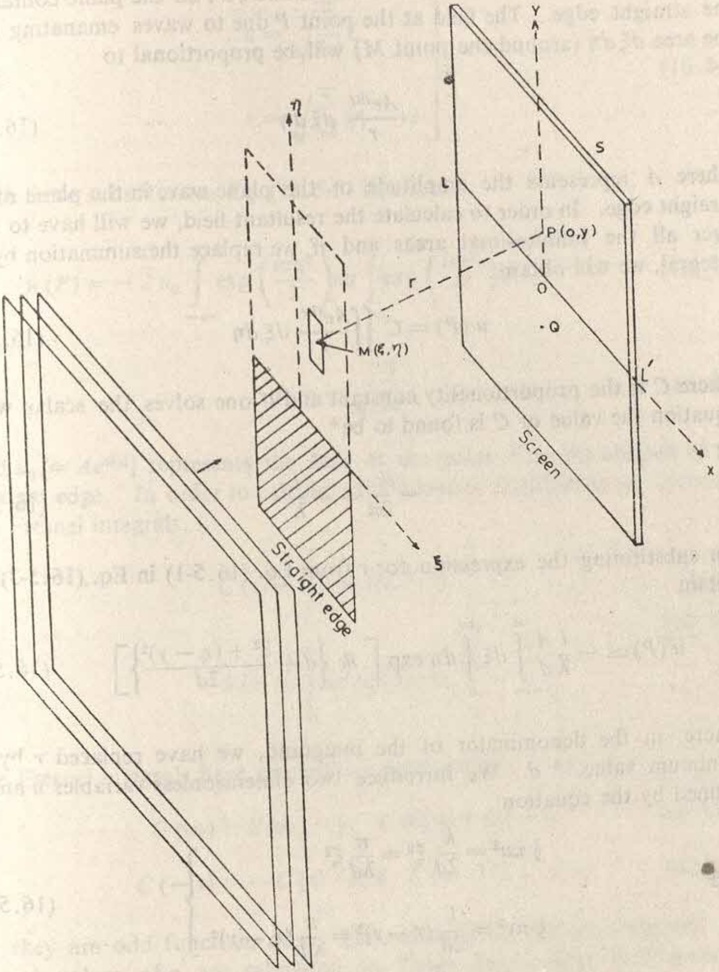


Fig. 16.8 Diffraction of a plane wave incident normally on a straight edge.

of the geometrical shadow. If the x and y coordinates of an arbitrary point M on the plane of the straight edge are denoted by ξ and η, then

$$r = MP = [\xi^2 + (\eta - y)^2 + d^2]^{1/2} = d\left[1 + \frac{\xi^2 + (\eta - y)^2}{d^2}\right]^{1/2}$$

$$\simeq d + \frac{\xi^2 + (\eta - y)^2}{2d} \qquad (16.5\text{-}1)$$

where d is the distance between the straight edge and the screen. The field at the point P can be approximately calculated by using Huygens' principle. We consider an infinitesimal area $d\xi\, d\eta$ on the plane containing the straight edge. The field at the point P due to waves emanating from the area $d\xi\, d\eta$ (around the point M) will be proportional to

$$\frac{Ae^{ikr}}{r}\, d\xi\, d\eta \tag{16.5-2}$$

where A represents the amplitude of the plane wave in the plane of the straight edge. In order to calculate the resultant field, we will have to sum over all the infinitesimal areas and if we replace the summation by an integral, we will obtain

$$u(P) = C \iint \frac{Ae^{ikr}}{r}\, d\xi\, d\eta \tag{16.5-3}$$

where C is the proportionality constant and if one solves the scalar wave equation the value of C is found to be*

$$-\frac{ik}{2\pi} = -\frac{i}{\lambda} \tag{16.5-4}$$

On substituting the expression for r from Eq. (16.5-1) in Eq. (16.5-3) we obtain

$$u(P) \simeq -\frac{i}{\lambda}\frac{A}{d} \int_{-\infty}^{\infty} d\xi \int_{0}^{\infty} d\eta \exp\left[ik \left\{ d + \frac{\xi^2 + (\eta - y)^2}{2d} \right\} \right] \tag{16.5-5}$$

where, in the denominator of the integrand, we have replaced r by its minimum value,** d. We introduce two dimensionless variables u and v defined by the equation

$$\left.\begin{aligned} \tfrac{1}{2}\pi u^2 &= \frac{k}{2d}\xi^2 = \frac{\pi}{\lambda d}\xi^2 \\ \text{and} \qquad \tfrac{1}{2}\pi v^2 &= \frac{k}{2d}(\eta - y)^2 = \frac{\pi}{\lambda d}(\eta - y)^2 \end{aligned}\right\} \tag{16.5-6}$$

---

*The value of the constant can also be determined by noting that in the absence of the straight edge $u(P)$ must be equal to Ae^{ikd}. Thus

$$C \iint_{-\infty}^{\infty} A \frac{e^{ikr}}{r} d\xi\, d\eta$$

must be equal to Ae^{ikd}. If we carry out the integration we indeed find that $C = -\frac{i}{\lambda}$ (see Eq. 16.5-16).

**This is justified because in carrying out the integration, only a small region around the point $r = d$ contributes; the contribution due to far-off points is small because of the rapid oscillations of the exponential term in the integrand (see also footnote on p. 375).

so that

and
$$\left.\begin{aligned} u &= \sqrt{\frac{2}{\lambda d}}\,\xi \\ v &= \sqrt{\frac{2}{\lambda d}}\,(\eta - y) \end{aligned}\right\} \qquad (16.5\text{-}6)$$

With these substitutions Eq. (16.5-5) becomes

$$u\,(P) = -\frac{i}{2}\,u_0 \int_{-\infty}^{\infty} \exp\left(\frac{i\pi u^2}{2}\right) du \int_{v_0}^{\infty} \exp\left(\frac{i\pi v^2}{2}\right) dv \qquad (16.5\text{-}7)$$

where

$$v_0 = -\sqrt{\frac{2}{\lambda d}}\,y \qquad (16.5\text{-}8)$$

and $u_0\,[= Ae^{ikd}]$ represents the field at the point P in the absence of the straight edge. In order to calculate the intensity distribution we introduce the Fresnel integrals

$$\left.\begin{aligned} C\,(\tau) &= \int_0^{\tau} \cos\left(\tfrac{1}{2}\pi v^2\right) dv \\ S\,(\tau) &= \int_0^{\tau} \sin\left(\tfrac{1}{2}\pi v^2\right) dv \end{aligned}\right\} \qquad (16.5\text{-}9)$$

The Fresnel integrals have the following properties

$$C\,(\infty) = S\,(\infty) = \tfrac{1}{2}; \quad C\,(0) = S\,(0) = 0 \qquad (16.5\text{-}10)$$

$$C\,(-\tau) = -C\,(\tau) \quad \text{and} \quad S\,(-\tau) = -S\,(\tau) \qquad (16.5\text{-}11)$$

i.e. they are odd functions of τ. The values of the Fresnel integrals for typical values of τ are tabulated in Table 16.1. Fig. 16.9 gives a geometrical representation of the Fresnel integrals and is known as the Cornu's spiral. The numbers written on the spiral are the values of τ. For example, as can be seen from the figure, for $\tau = 1.5$, $C\,(\tau) \simeq 3.53$ and $S\,(\tau) \simeq 0.445$. Since the functions $\cos\frac{\pi}{2}\,v^2$ and $\sin\frac{\pi}{2}\,v^2$ are even functions of v, we have

$$\begin{aligned} \int_{-\infty}^{\infty} \exp\left[\frac{i\pi u^2}{2}\right] du &= 2\left[\int_0^{\infty} \cos\left(\frac{\pi}{2}\,u^2\right) du + i \int_0^{\infty} \sin\left(\frac{\pi}{2}\,u^2\right) du\right] \\ &= 2\left[\tfrac{1}{2} + i\,\tfrac{1}{2}\right] = (1 + i) \end{aligned} \qquad (16.5\text{-}12)$$

TABLE 16.1

Table of Fresnel Integrals

$$C(\tau) = \int_0^{\tau} \cos\left(\frac{\pi}{2}\ \nu^2\right) d\nu;\ \ S(\tau) = \int_0^{\tau} \sin\left(\frac{\pi}{2}\ \nu^2\right) d\nu$$

τ	$C(\tau)$	$S(\tau)$	τ	$C(\tau)$	$S(\tau)$
0.0	0.00000	0.00000	2.6	0.38894	0.54999
0.2	0.19992	0.00419	2.8	0.46749	0.39153
0.4	0.39748	0.03336	3.0	0.60572	0.49631
0.6	0.58110	0.11054	3.2	0.46632	0.59335
0.8	0.72284	0.24934	3.4	0.43849	0.42965
1.0	0.77989	0.43826	3.6	0.58795	0.49231
1.2	0.71544	0.62340	3.8	0.44809	0.56562
1.4	0.54310	0.71353	4.0	0.49843	0.42052
1.6	0.36546	0.63889	4.2	0.54172	0.56320
1.8	0.33363	0.45094	4.4	0.43833	0.46227
2.0	0.48825	0.34342	4.6	0.56724	0.51619
2.2	0.63629	0.45570	4.8	0.43380	0.49675
2.4	0.55496	0.61969	5.0	0.56363	0.49919

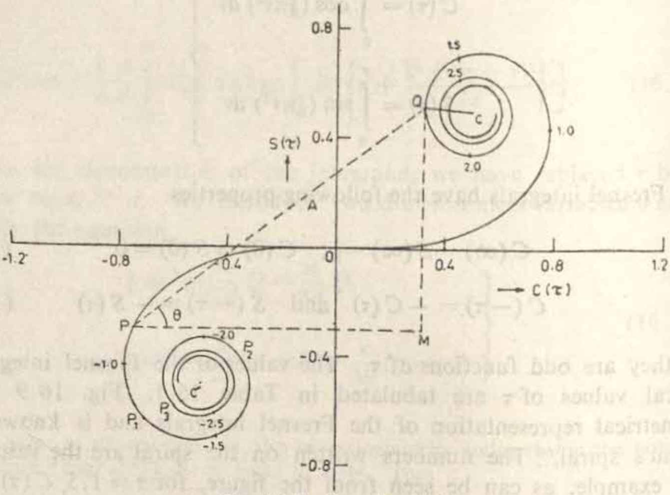


Fig. 16.9 Cornu's spiral.

Further

$$\int_{\nu_0}^{\infty} \exp\left(\frac{i\pi\nu^2}{2}\right) d\nu = \left[\int_0^{\infty} \cos\left(\frac{\pi}{2}\,\nu^2\right) d\nu - \int_0^{\nu_0} \cos\left(\frac{\pi}{2}\,\nu^2\right) d\nu\right]$$
$$+\, i\left[\int_0^{\infty} \sin\left(\frac{\pi}{2}\,\nu^2\right) d\nu - \int_0^{\nu_0} \sin\left(\frac{\pi}{2}\,\nu^2\right) d\nu\right] \qquad (16.5\text{-}13)$$

Thus, if we use Eq. (16.5-10) we would obtain

$$\int_{\nu_0}^{\infty} \exp\left[\frac{i\pi\nu^2}{2}\right] d\nu = [\tfrac{1}{2} - C(\nu_0)] + i[\tfrac{1}{2} - S(\nu_0)] \qquad (16.5\text{-}14)$$

Substituting in Eq. (16.5-7), we obtain

$$\begin{aligned} u(P) &= -\frac{i}{2} u_0 (1+i)[\{\tfrac{1}{2} - C(\nu_0)\} + i\{\tfrac{1}{2} - S(\nu_0)\}] \\ &= \frac{1-i}{2} u_0 [\{\tfrac{1}{2} - C(\nu_0)\} + i\{\tfrac{1}{2} - S(\nu_0)\}] \qquad (16.5\text{-}15) \end{aligned}$$

It is of interest to note that a large value of y corresponds to a point which is very far away from the edge of the geometrical shadow. For such a point ν_0 would tend to $-\infty$ [see Eq. (16.5-8)] and we would obtain

$$\begin{aligned} u(P) &= \frac{1-i}{2} u_0 [(\tfrac{1}{2} + \tfrac{1}{2}) + i(\tfrac{1}{2} + \tfrac{1}{2})] \\ &= u_0 \qquad (16.5\text{-}16) \end{aligned}$$

which implies the fact that the amplitude at such a point should be the same as that in the absence of the edge. This also justifies the value of the constant given by Eq. (16.5-4).

The intensity distribution corresponding to Eq. (16.5-15) would be given by

$$I(P) = \tfrac{1}{2} I_0 [\{\tfrac{1}{2} - C(\nu_0)\}^2 + \{\tfrac{1}{2} - S(\nu_0)\}^2] \qquad (16.5\text{-}17)$$

If the point P is such that it lies on the edge of the geometrical shadow (i.e. on the line LL'—see Fig. 16.8) then $y = 0$ and hence $\nu_0 = 0$; thus

$$I = \tfrac{1}{2} I_0 [\tfrac{1}{4} + \tfrac{1}{4}] = \tfrac{1}{4} I_0 \qquad (16.5\text{-}18)$$

where we have used the fact that $C(0) = S(0) = 0$. Thus the intensity on the edge of the geometrical shadow is 1/4th of the intensity that would have been in the absence of the edge (see Eq. 16.4-3). On the other hand, when the point P is deep inside the geometrical shadow (i.e. when $y \to -\infty$ and hence $\nu_0 \to \infty$), we obtain $C(\nu_0) = S(\nu_0) \to \tfrac{1}{2}$ giving

$$u(P) \to 0$$

as it should indeed be.

In order to determine the field at an arbitrary point P, we may use Table 16.1 to calculate the RHS of Eq. (16.5-17). The intensity variation is plotted in Fig. 16.10 from which one can make the following observations:

(i) Fig. 16.10 represents a universal curve; i.e. for given values of λ and d, one simply has to calculate ν_0 as the observation point moves along

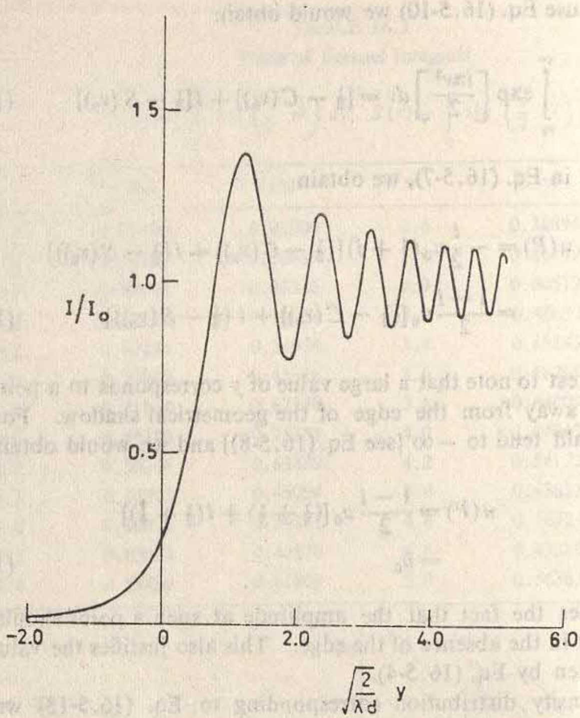


Fig. 16.10 Intensity distribution in the straight edge diffraction pattern.

the y-axis. For example, for $\lambda = 6\times10^{-5}$ cm, $d = 120$ cm

$$y = -\sqrt{\frac{\lambda d}{2}}\, v_0 = -0.06\, v_0 \text{ cm} \tag{16.5-19}$$

Since the first three maxima correspond to $v_0 \simeq -1.22$, -2.34 and -3.08; the positions of the maxima in the actual diffraction pattern occur at $y \simeq 0.732$, 1.404 and 1.848 mm respectively. The intensities at these maxima are approximately $1.37\, I_0$, $1.20\, I_0$ and $1.15\, I_0$ respectively. The first three minima occur at $v_0 \simeq -1.87$, -2.74 and -3.39 and the corresponding intensities are approximately $0.778\, I_0$, $0.843\, I_0$ and $0.872\, I_0$ respectively. These results may be compared with those obtained in Sec. 16.4.

(ii) As we go inside the geometrical shadow the intensity monotonically decreases to zero.

(iii) One could have also studied the intensity variation directly from the Cornu's spiral (see Fig. 16.9). This is due to the fact that associated with the Cronu's spiral, we have the following interesting property: let us write

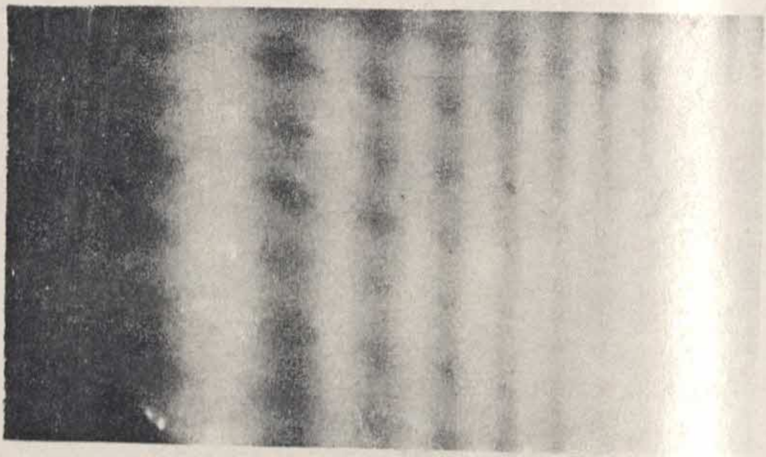


Plate 16 The straight edge diffraction pattern. (Photograph courtesy of Dr. K. Thyagarajan.)

$$[C(\tau_2)-C(\tau_1)]+i[S(\tau_2)-S(\tau_1)]\equiv Ae^{i\theta} \qquad (16.5\text{-}20)$$

Thus

$$C(\tau_2)-C(\tau_1)=A\cos\theta$$

and

$$S(\tau_2)-S(\tau_1)=A\sin\theta$$

Let the points P and Q on the Cornu's spiral (see Fig. 16.9) correspond to $\tau=\tau_1$ and $\tau=\tau_2$ respectively. It is obvious that

and

$$\left.\begin{aligned} PM&=C(\tau_2)-C(\tau_1)=A\cos\theta \\ QM&=S(\tau_2)-S(\tau_1)=A\sin\theta \end{aligned}\right\} \qquad (16.5\text{-}21)$$

Thus the length of the line joining the points P and Q will be A and the angle that the line makes with the abcissa will be θ.

In order to use the Cornu's spiral we rewrite Eq. (16.5-15):

$$u=\frac{1-i}{2}u_0[\{\tfrac{1}{2}-C(\nu_0)\}+i\{\tfrac{1}{2}-S(\nu_0)\}] \qquad (16.5\text{-}22)$$

Let us first consider a point of observation Q in the geometrical shadow region. Consequently ν_0 will be positive. Let the point Q on the spiral (see Fig. 16.9) correspond to $\tau=\nu_0$. Since the point C in the curve corresponds to $\tau=\infty$, we have

$$\{\tfrac{1}{2}-C(\nu_0)\}+i\{\tfrac{1}{2}-S(\nu_0)\}=(QC)e^{i\Psi} \qquad (16.5\text{-}23)$$

where Ψ is the angle that QC makes with the abcissa [see Eq. (16.5-20)]. Thus,

$$u(Q)=\frac{1-i}{2}(QC)e^{i\Psi}u_0$$

or

$$I(Q)=\tfrac{1}{4}(QC)^2I_0 \qquad (16.5\text{-}24)$$

We can easily see that as the point of observation moves into the shadow region, the value of ν_0 increases. Thus the point Q keeps on moving on the spiral towards the point C and the length QC decreases uniformly. Hence in the shadow region the intensity uniformly decreases to zero (see Fig. 16.10 and Plate 16).

As we move away from the edge of the geometrical shadow to the illuminated region, the value of ν_0 becomes negative and the corresponding point P (on the Cornu's spiral) lies in the third quadrant as shown in Fig. 16.9. The intensity is again given by

$$I(P)=\tfrac{1}{4}(PC)^2I_0 \qquad (16.5\text{-}25)$$

As the value of ν_0 becomes more and more negative, the length PC keeps on increasing till the point P reaches P_1 which corresponds to $\nu_0\simeq-1.22$.

The intensity at this point is maximum and the length $P_1C \simeq 2.34$. Thus, the corresponding intensity is

$$I(P_1) \simeq \tfrac{1}{4}(2.34)^2 I_0 \simeq 1.37\, I_0 \qquad (16.5\text{-}26)$$

As the value of v_0 becomes further negative the length PC starts decreasing till it reaches the point P_2. Thus, the intensity keeps on oscillating with decreasing amplitude about I_0 as we move more and more into the illuminated region (see Fig. 16.10 and Plate 16).

16.6 Diffraction of a Plane Wave by a Long Narrow Slit and Transition to the Fraunhofer Region

We next consider a plane wave incident normally on a long narrow slit (of width b) as shown in Fig. 16.11. We wish to calculate the intensity

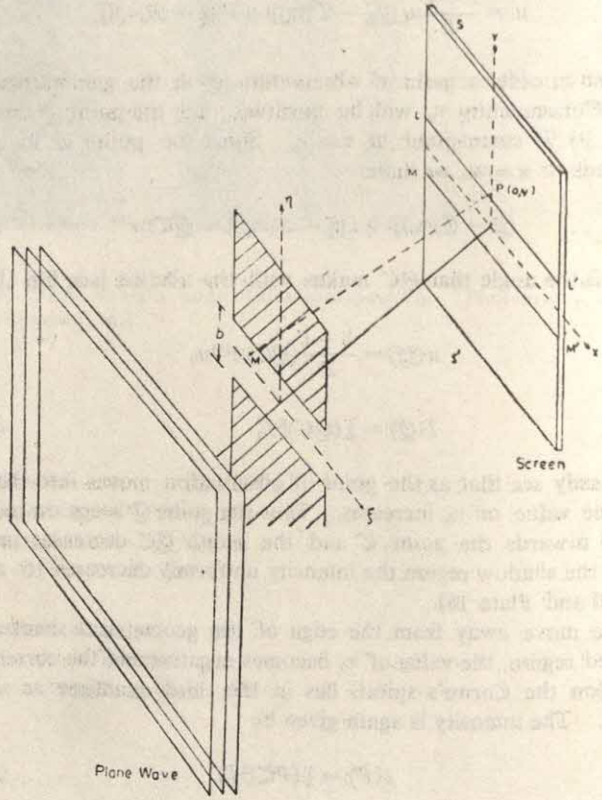


Fig. 16.11 Diffraction of a plane wave incident normally on a long narrow slit.

distribution at an arbitrary point P on the screen SS'. The lines LL' and MM' represent the edges of the geometrical shadow. Once again, there will be no variation of the intensity along the x-axis and we may (without any loss of generality) assume the coordinates of the point P to be $(0, y)$. The field at the point P will again be given by Eq. (16.5-5) except that the limits of the η-integral will be $-b/2$ and $+b/2$ (we are assuming the origin to be at the centre of the slit). Thus

$$u(P) \simeq -\frac{iA}{\lambda d}\int_{-\infty}^{+\infty} d\xi \int_{-b/2}^{+b/2} d\eta \exp\left[ik\left\{d+\frac{\xi^2+(\eta-y)^2}{2d}\right\}\right] \qquad (16.6\text{-}1)$$

Carrying out manipulations similar to that in the previous section we obtain

$$u(P) = -\frac{i}{2}u_0 \int_{-\infty}^{\infty} \exp\left[\frac{i\pi u^2}{2}\right] du \int_{-(v_2+v_1)}^{-(v_2-v_1)} \exp\left[\frac{i\pi v^2}{2}\right] dv \qquad (16.6\text{-}2)$$

where

$$u = \sqrt{\frac{2}{\lambda d}}\,\xi, \quad v = \sqrt{\frac{2}{\lambda d}}\,(\eta - y)$$

and

$$v_1 = \sqrt{\frac{2}{\lambda d}}\,\frac{b}{2} \quad \text{and} \quad v_2 = \sqrt{\frac{2}{\lambda d}}\,y \qquad (16.6\text{-}3)$$

Using Eq. (16.5-12) we obtain

$$u(P) = -\frac{i}{2}u_0(1+i)\left[\int_0^{-(v_2-v_1)}\left[\cos\left(\frac{\pi}{2}v^2\right)+i\sin\left(\frac{\pi}{2}v^2\right)\right]dv - \int_0^{-(v_2+v_1)}\left[\cos\left(\frac{\pi}{2}v^2\right)+i\sin\left(\frac{\pi}{2}v^2\right)\right]dv\right]$$

or

$$u(P) = \frac{(1-i)}{2}u_0[\{C(v_2+v_1) - C(v_2-v_1)\} + i\{S(v_2+v_1) - S(v_2-v_1)\}] \qquad (16.6\text{-}4)$$

where we have used Eq. (16.5-11). Thus the intensity distribution would be

$$I(P) = \tfrac{1}{2}I_0[\{C(v_2+v_1) - C(v_2-v_1)\}^2 + \{S(v_2+v_1) - S(v_2-v_1)\}^2] \qquad (16.6\text{-}5)$$

For a given system λ, d and b are known which determine v_1; e.g. for $\lambda = 5\times10^{-5}$ cm, $d = 100$ cm and $b = 0.1$ cm, one obtains $v_1 = 2.0$; further, as y varies on the screen, the quantity v_2 also changes. In Figs. 16.12 (a), (b), (c) and (d) we have plotted the intensity variation as a

On the other hand, for small values of v_1 (i.e. when the observation screen is far away from the aperture) the diffraction pattern is essentially of the Fraunhofer type. In order to show this explicitly we notice that

$$v_2 = \sqrt{\frac{2}{\lambda d}}\, y = \sqrt{\frac{2d}{\lambda}}\, \frac{y}{d} \simeq \sqrt{\frac{2d}{\lambda}}\, \theta \tag{16.6-6}$$

where θ represents the angle of diffraction (see Fig. 16.13). Clearly, in the Fraunhofer region since d is very large, the value of v_2 will also be very large and thus we must look for expressions of the Fresnel integrals in the limit of $v \to \infty$. Now, we may write

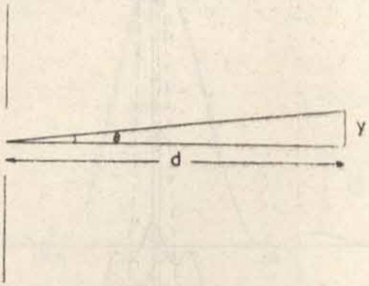


Fig. 16.13 In the Fraunhofer region, *d* is very large.

$$\begin{aligned}
C(v) &= \int_0^v \cos\frac{\pi}{2} v^2\, dv = \int_0^\infty \cos\frac{\pi}{2} v^2\, dv - \int_v^\infty \cos\frac{\pi}{2} v^2 dv \\
&= \tfrac{1}{2} - \int_v^\infty \frac{1}{\pi v} \cos\left(\frac{\pi}{2} v^2\right) \pi v\, dv \\
&= \tfrac{1}{2} - \frac{1}{\pi v} \sin\left(\frac{\pi}{2} v^2\right)\Bigg|_v^\infty + \int_v^\infty \frac{1}{\pi v^2} \sin\left(\frac{\pi}{2} v^2\right) dv \\
&\cong \tfrac{1}{2} + \frac{1}{\pi v} \sin\frac{\pi v^2}{2}
\end{aligned} \tag{16.6-7}$$

where we have neglected terms which would be of order $1/v^3$. Similarly

$$S(v) \simeq \tfrac{1}{2} - \frac{1}{\pi v} \cos\left(\frac{\pi}{2} v^2\right) \tag{16.6-8}$$

Since v_2 is large and v_1 is small we have

$$\begin{aligned}
C(v_2 + v_1) - C(v_2 - v_1) \simeq &\left[\tfrac{1}{2} + \frac{1}{\pi v_2} \sin\frac{\pi}{2} (v_2 + v_1)^2\right] \\
&- \left[\tfrac{1}{2} + \frac{1}{\pi v_2} \sin\frac{\pi}{2} (v_2 - v_1)^2\right]
\end{aligned}$$

$$\simeq \frac{2}{\pi v_2} \cos \frac{\pi}{2} (v_2^2 + v_1^2) \sin \pi v_1 v_2 \tag{16.6-9}$$

Similarly

$$S(v_2 + v_1) - S(v_2 - v_1) \simeq \frac{2}{\pi v_2} \sin \frac{\pi}{2} (v_2^2 + v_1^2) \sin \pi v_1 v_2 \tag{16.6-10}$$

Thus in the Fraunhofer limit, Eq. (16.6-5) becomes

$$\begin{aligned} I(P) &= \tfrac{1}{2} I_0 \left[\frac{4}{\pi^2 v_2^2} \sin^2 (\pi v_1 v_2) \right] \\ &= I_{00} \frac{\sin^2 \beta}{\beta^2} \end{aligned} \tag{16.6-11}$$

where

$$I_{00} = 2 I_0 v_1^2$$

and

$$\beta = \pi v_1 v_2 = \frac{\pi}{\lambda d} b y \simeq \frac{\pi b}{\lambda} \theta \tag{16.6-12}$$

which shows that the intensity distribution is indeed of the Fraunhofer type. In Fig. 16.12 the dotted curves correspond to Eq. (16.6-11) and one can see that the intensity distribution is almost of the Fraunhofer type for $V_1 \lesssim 0.5$.

16.7 Fresnel Diffraction by a Circular Aperture

In Sections 16.5 and 16.6 we had considered the diffraction of plane waves incident normally on a straight edge and on a long narrow slit respectively. In this section we will first obtain a general formula for diffraction of spherical waves by an aperture of arbitrary shape, which will be followed by a discussion of the diffraction pattern produced by a circular aperture.

Let us consider an aperture A illuminated by a point source P'. In order to calculate the field at an arbitrary point P, we again use Huygens' principle. We consider an infinitesimal area ΔS on the aperture A. The field at the point P due to the secondary wavelets emanating from the area ΔS would be proportional to $\frac{1}{r} \Delta S\, e^{ikr}$. But the field at the area ΔS due to a point source at P' will be $\frac{A}{\sigma} e^{ik\sigma}$, where the distances r and σ are as shown in Fig. 16.14. Thus, the field at the point P would be proportional to

$$A \sum \frac{e^{ik(r+\sigma)}}{r\sigma} \Delta S \tag{16.7-1}$$

where the summation is carried over the entire aperture. Therefore, the field at the point P will be given by

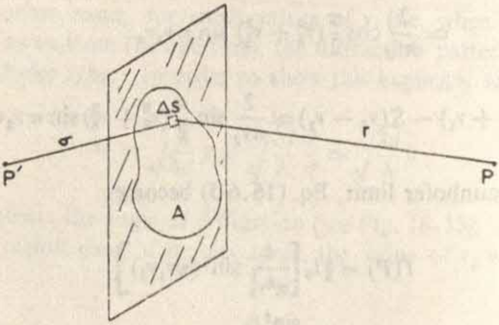


Fig. 16.14 Diffraction of a spherical wave (emanating from the point P') by an aperture.

$$u(P) = -\frac{iA}{\lambda}\iint_A \frac{e^{ik(r+\sigma)}}{r\sigma}\, dS \tag{16.7-2}$$

where we have used Eq. (16.5-4) and the integration is carried over the area of the aperture.

We next assume the aperture to be circular of radius a, illuminated by a point source P' which is assumed to be on the axis [see Fig. 16.15 (a)];

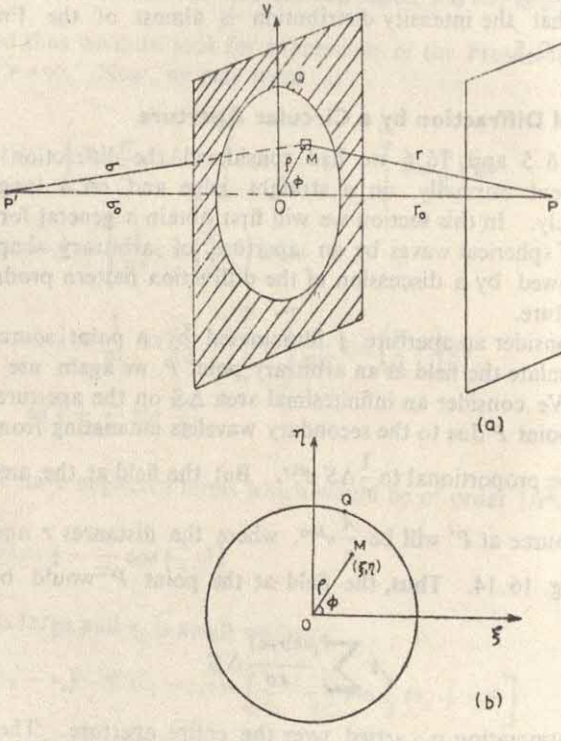


Fig. 16.15 (a) Diffraction by a circular aperture of a spherical wave emanating from a point on the axis. (b) The coordinate system.

the axis is defined to be a line passing through the centre of the circle and perpendicular to its plane. It is difficult to calculate the intensity distribution on a plane normal to the axis and, therefore, for the sake of mathematical simplicity we will calculate the variation of intensity only along the axis. Obviously, it will be more convenient to use the cylindrical system of coordinates. In this system, the coordinates of an arbitrary point M on the aperture will be (ρ, φ) where ρ is the distance of the point M from the centre O and φ is the angle that OM makes with the x-axis [see Fig. 16.15 (a)] and dS will be $\rho d\rho\, d\varphi$. Thus

$$u(P) = -\frac{iA}{\lambda}\int_0^{2\pi}\int_0^{a}\frac{e^{ik(r+\sigma)}}{k\sigma}\,\rho d\rho\, d\varphi \tag{16.7-3}$$

Clearly, for a given value of ρ, the values of r and σ do not change as φ goes from 0 to 2π; thus,

$$u(P) = -\frac{2\pi iA}{\lambda}\int_0^{a}\frac{e^{ik(r+\sigma)}}{r\sigma}\,\rho\, d\rho \tag{16.7-4}$$

Since the angles MOP' and MOP are right angles [see Fig. 16.15(a)], we will have

$$\sigma^2 = \sigma_0^2 + \rho^2 \quad \text{and} \quad r^2 = r_0^2 + \rho^2 \tag{16.7-5}$$

Therefore

$$\sigma\, d\sigma = \rho\, d\rho \qquad \text{and} \quad r\, dr = \rho\, d\rho \tag{16.7-6}$$

or,

$$d(\sigma + r) = \left(\frac{1}{\sigma} + \frac{1}{r}\right)\rho\, d\rho \tag{16.7-7}$$

i.e.

$$\rho\, d\rho = \frac{\sigma r}{\sigma + r}\, d(\sigma + r) \tag{16.7-8}$$

Substituting the above value of $\rho\, d\rho$ in Eq. (16.7-4) we obtain

$$u(P) = -\frac{2\pi iA}{\lambda}\int\frac{e^{ik(\sigma+r)}}{\sigma + r}\, d(\sigma + r) \tag{16.7-9}$$

For a small aperture, the quantities r and σ do not vary appreciably over the aperture and as such very little error will be involved if we replace*

---

*It may be worthwhile to mention that although the quantities r and σ do not vary appreciably over the aperture, the factor exp $[ik\,(r + \sigma)]$ does have a very rapid variation. For example, for $\lambda = 6 \times 10^{-5}$ cm, the factor cos kr becomes

$$\cos\left(\frac{\pi}{3}\, 10^5\, r\right)$$

As the value of r is changed from say 60 cm to 60.00002 cm, the cosine factor will change from +1 to −0.5. This shows the rapidity with which the exponential factor will vary in the domain of integration.

r and σ in the denominator of the integrand by r_0 and σ_0. Thus we may write

$$u(P) \simeq -\frac{2\pi i}{\lambda}\frac{A}{r_0+\sigma_0}\int_{\zeta_1}^{\zeta_2} e^{ik\zeta}\, d\zeta \tag{16.7-10}$$

where

$$\zeta = r + \sigma$$

and

$$\zeta_1 = r_0 + \sigma_0;\ \zeta_2 = (r_0^2 + a^2)^{1/2} + (\sigma_0^2 + a^2)^{1/2}$$

Carrying out the above integration we obtain,

$$\begin{aligned} u(P) &= \frac{Ae^{ik(r_0+\sigma_0)}}{r_0+\sigma_0}[1 - e^{ik(\zeta_2-\zeta_1)}] \\ &= u_0(P)\,[1 - e^{ikp\lambda/2}] = u_0(P)\,[1 - e^{ip\pi}] \end{aligned} \tag{16.7-11}$$

where

$$u_0(P) = A\frac{e^{ik(r_0+\sigma_0)}}{r_0+\sigma_0} \tag{16.7-12}$$

represents the field at the point P in the absence of the aperture and

$$\frac{p\lambda}{2} \equiv \zeta_2 - \zeta_1 = [(r_0^2 + a^2)^{1/2} + (\sigma_0^2 + a^2)^{1/2}] - (r_0 + \sigma_0) \tag{16.7-13}$$

Thus,

$$I(P) = 4I_0 \sin^2\frac{p\pi}{2} \tag{16.7-14}$$

which shows that the intensity is zero or maximum when p is an even or odd integer, i.e. $(P'Q + QP) - (P'O + O'P)$ is an even or odd multiple of $\lambda/2$. This can be understood physically by using the concept of Fresnel half-period zones; the n^{th} circle is defined by the following relation (see Sec. 16.2)

$$(P'M_n + M_nP) - (P'O + OP) = n\lambda \tag{16.7-15}$$

The field produced at the point P by the n^{th} zone will be out of phase with the field produced by the $(n-1)^{\text{th}}$ zone. Consequently if the aperture contains an even number of half-period zones, the intensity at the point P will be negligibly small and conversely, if the circular aperture contains an odd number of zones, the intensity at the point P will be maximum, which is consistent with the results obtained in Sec. 16.2.

16.8 Diffraction by a Circular Disc

We next consider the diffraction pattern produced by an opaque disc of radius a (see Fig. 16.4). Once again we will assume that the point source as well as the observation point lie on the axis of the disc. Eq. (16.7-2) tells us that in order to calculate the field we have to carry out an integration over the open region of the aperture. Obviously, if $u_1(P)$ and $u_2(P)$ respectively represent the fields at the point P due to a circular aperture and an opaque disc (of the same radius), then

$$u_1(P) + u_2(P) = u_0(P) \tag{16.8-1}$$

where $u_0(P)$ represents the field in the absence of any aperture. [Eq. (16.8-1) is known as the Babinet's principle]. Thus,

$$\begin{aligned} u_2(P) &= u_0(P) - u_1(P) \\ &= u_0(P) - u_0(P)\,[1 - e^{ip\pi}] \\ &= u_0(P)\, e^{ip\pi} \end{aligned} \tag{16.8-2}$$

where, for $u_1(P)$ we have used Eq. (16.7-11). Thus the intensity at the point P on the axis of a circular disc would be

$$\begin{aligned} I_2(P) &= |\, u_2(P)\, |^2 \\ &= I_0(P) \end{aligned} \tag{16.8-3}$$

which gives us the remarkable result that the intensity at a point on the axis of an opaque disc is equal to the intensity at the point in the absence of the disc ! Notice that this result is true only as long as Eq. (16.7-10) is true, i.e. only as long as the disc is so small that the variation in $(\sigma + r)$ is negligibly small over the disc.

16.9 Fraunhofer Diffraction by an Aperture

We now use the method developed in Sec. 16.5 and 16.6 to study the Fraunhofer diffraction pattern produced by a plane wave incident normally on an aperture as shown in Fig. 16.16. The field at the point P will again be given by Eq. (16.5-3):

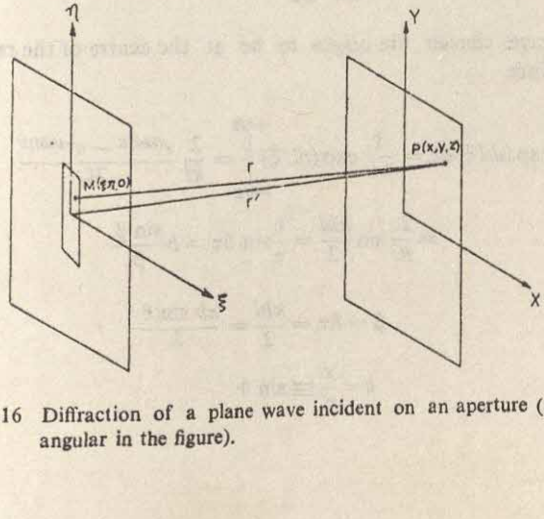


Fig. 16.16 Diffraction of a plane wave incident on an aperture (shown rectangular in the figure).

$$u(P) = -\frac{iA}{\lambda}\iint \frac{e^{ikr}}{r}\,d\xi\,d\eta \tag{16.9-1}$$

where $r = PM$, M representing an arbitrary point on the aperture. If the coordinates of the point P and M are (x, y, z) and $(\xi, \eta, 0)$ then

$$\begin{aligned} r = PM &= [(x-\xi)^2 + (y-\eta)^2 + z^2]^{1/2} \\ &= r'\left[1 - \frac{2(x\xi + y\eta)}{r'^2} + \frac{\xi^2 + \eta^2}{r'^2}\right]^{1/2} \\ &\cong r' - \frac{(x\xi + y\eta)}{r'} \end{aligned} \tag{16.9-2}$$

where $r' = \sqrt{x^2 + y^2 + z^2}$ represents the distance of the point P from a conveniently chosen origin O. In writing Eq. (16.9-2) we have neglected the terms of order $1/r'^2$; this corresponds to Fraunhofer diffraction. Substituting for r from Eq. (16.9-2) in Eq. (16.9-1) we obtain

$$u(P) \simeq C\iint e^{-ik(l\xi + m\eta)}\,d\xi\,d\eta \tag{16.9-3}$$

where

$$C = -\frac{iA\,e^{ikr'}}{\lambda r'}$$

and $l = x/r'$ and $m = y/r'$ represent the direction cosines of the diffracted ray.

16.9.1 Rectangular Aperture

For a rectangular aperture (of dimensions a and b) we obtain

$$u(P) = C\int_{-b/2}^{+b/2}\int_{-a/2}^{+a/2} e^{ik(l\xi + m\eta)}\,d\xi\,d\eta \tag{16.9-4}$$

where we have chosen the origin to be at the centre of the rectangular aperture. Since

$$\begin{aligned} \int_{-b/2}^{b/2} \exp(ikl\xi)\,d\xi &= \frac{1}{ikl}\exp(ikl\xi)\Big|_{-b/2}^{+b/2} = \frac{2}{kl}\,\frac{e^{ikbl/2} - e^{-ikbl/2}}{2i} \\ &= \frac{2}{kl}\sin\frac{kbl}{2} = \frac{1}{\tau}\sin b\tau = b\,\frac{\sin\beta}{\beta} \end{aligned} \tag{16.9-5}$$

where

$$\beta = b\tau = \frac{kbl}{2} = \frac{\pi b\sin\theta}{\lambda} \tag{16.9-6}$$

$$l = \frac{x}{r'} \equiv \sin\theta$$

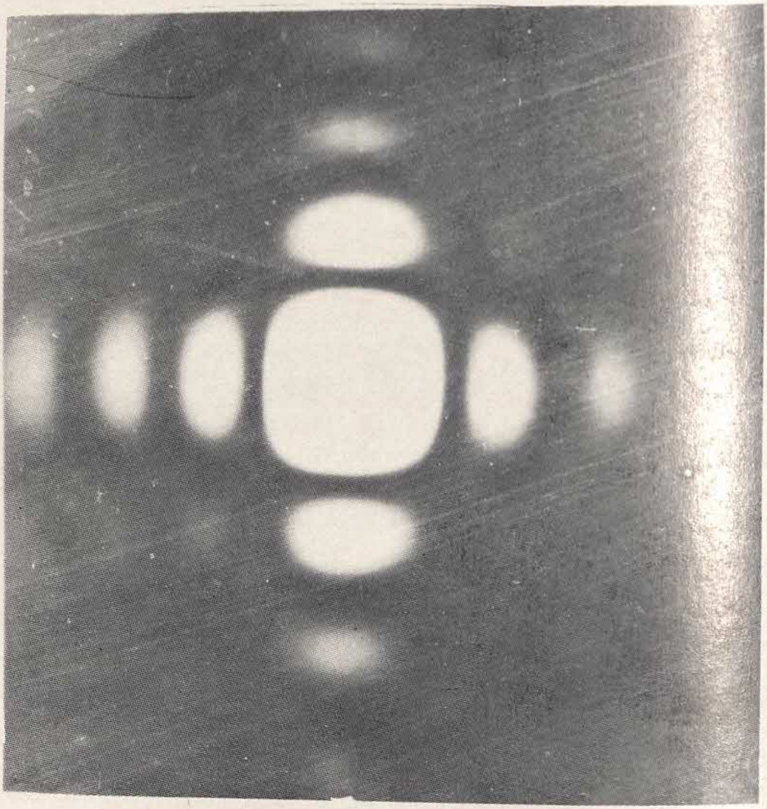


Plate 17 Fraunhofer diffraction pattern of a square aperture. (Photograph courtesy Mr. K. K. Gupta.)

Similarly

$$\int_{-a/2}^{a/2} e^{ikm\eta}\, d\eta = \frac{1}{\zeta} \sin a\zeta = a \frac{\sin \gamma}{\gamma} \tag{16.9-7}$$

where

$$\gamma = a\tau = \frac{kam}{2} = \frac{\pi a \sin \varphi}{\lambda} \tag{16.9-8}$$

$$m = \frac{y}{r'} \equiv \sin \varphi$$

Thus

$$u(P) = C \frac{\sin a\zeta}{\zeta} \frac{\sin b\tau}{\tau} \tag{16.9-9}$$

which can be rewritten in the form

$$u(P) = u_0 \frac{\sin \gamma}{\gamma} \frac{\sin \beta}{\beta} \tag{16.9-10}$$

where

$$u_0 = C ab \tag{16.9-11}$$

represents the amplitude at the axial point (i.e., $\theta = 0$ and $\varphi = 0$). The corresponding intensity distribution would be

$$I(P) = I_0 \frac{\sin^2 \gamma}{\gamma^2} \frac{\sin^2 \beta}{\beta^2} \tag{16.9-12}$$

which represents the Fraunhofer diffraction pattern by a rectangular aperture. It may be metioned that the angles θ and ϕ represent the diffraction angles along the x- and y-axes respectively. The intensity distribution due to a square aperture ($a = b$) is shown in Plate 17. For the case of a long narrow slit (i.e. for $a \rightarrow \infty$), the function $\frac{\sin a\zeta}{\zeta}$ becomes very sharply peaked around $\zeta = 0$ (i.e. $\varphi = 0$). Since, $\zeta = 0$ implies $y = 0$, the intensity appears only along the x-axis and there is no diffraction along the y-axis; (see Fig. 15.6).

16.9.2 Circular Aperture

For a circular aperture we choose the cylindrical coordinates [see Fig. 16.15(b)]

$$\xi = \rho \cos \varphi \quad \text{and} \quad \eta = \rho \sin \varphi \tag{16.9-13}$$

Further, because of the circular symmetry of the system the diffraction pattern will be of the form of concentric circular rings with their centres at the point O' (see Fig. 16.17). Consequently, we may calculate the intensity distribution only along the x-axis (i.e. at points for which $y = 0$). Thus

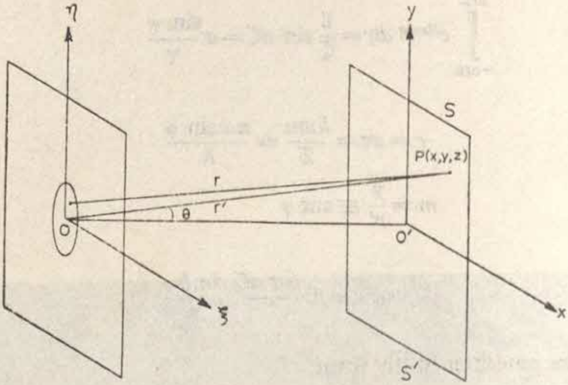


Fig. 16.17 Diffraction of a plane wave incident on a circular aperture.

$$m = 0 \text{ and } l = \frac{x}{r'} = \sin\theta \tag{16.9-14}$$

and Eq. (16.9-3) becomes

$$u(P) = C \int_0^a \int_0^{2\pi} e^{-ik\rho \sin\theta \cos\varphi}\, \rho\, d\rho\, d\varphi \tag{16.9-15}$$

or,

$$u(P) = \frac{C}{(k \sin\theta)^2} \int_0^{ka \sin\theta} \zeta\, d\zeta \int e^{-i\zeta \cos\varphi}\, d\varphi$$

$$= \frac{2\pi C}{(k \sin\theta)^2} \int_0^{ka \sin\theta} \zeta J_0(\zeta)\, d\zeta \tag{16.9-16}$$

where $\zeta = k\rho \sin\theta$ and use has been made of the following well known relation:

$$J_0(\zeta) = \frac{1}{2\pi} \int_0^{2\pi} \exp[\pm \xi \cos\varphi]\, d\varphi \tag{16.9-17}$$

If we further use the relation

$$\frac{d}{d\zeta}[\zeta J_1(\zeta)] = \zeta J_0(\zeta) \tag{16.9-18}$$

then Eq. (16.9-16) becomes

$$u(P) = \frac{2\pi C}{(k \sin\theta)^2} [\zeta J_1(\zeta)] \Big|_0^{ka \sin\theta}$$

$$= u_0 \left[\frac{2J_1(v)}{v}\right]$$

where $v = ka \sin \theta$. Thus the intensity distribution would be given by

$$I(P) = I_0 \left[\frac{2J_1(v)}{v}\right]^2 \tag{16 9-19}$$

This is the famous Airy pattern which has been discussed in Sec. 15.3.

We have already mentioned that the diffraction pattern (in the plane SS') will consist of concentric rings with their centres at the point O' (see Fig. 16.17). If $F(r)$ represents the fractional energy contained in a circle of radius r then

$$F(r) = \frac{\int_0^r I(\sigma)\, 2\pi\sigma d\sigma}{\int_0^\infty I(\sigma)\, 2\pi\sigma d\sigma} \tag{16.9-20}$$

where $I(\sigma)\, 2\pi\sigma d\sigma$ would be proportional to the energy contained in the annular region whose radii lie between σ and $\sigma + d\sigma$. Clearly

$$\frac{\sigma}{r'} = \sin \theta \tag{16.9-21}$$

Since $v = ka \sin \theta$, we obtain

$$\sigma = \frac{r'}{ka}\, v \tag{16.9-22}$$

Thus Eq. (16.9-20) becomes

$$F(r) = \frac{\int_0^v \left[\frac{2J_1(v)}{v}\right]^2 v dv}{\int_0^\infty \left[\frac{2J_1(v)}{v}\right]^2 v dv} \tag{16.9-23}$$

where we have used Eq. (16.9-19) for the intensity distribution. Now*,

$$\begin{aligned} \frac{J_1^2(v)}{v} &= J_1(v)\left[J_0(v) - \frac{dJ_1(v)}{dv}\right] \\ &= -\left[J_0(v)\frac{dJ_0(v)}{dv} + J_1(v)\frac{dJ_1(v)}{dv}\right] \\ &= -\tfrac{1}{2}\frac{d}{dv}[J_0^2(v) + J_1^2(v)] \end{aligned} \tag{16.9-24}$$

---

*See any book on mathematical physics for the recurence relations of Bessel functions.

Thus

$$F(r) = \frac{J_0^2(v) + J_1^2(v)\Big|_0^v}{J_0^2(v) + J_1^2(v)\Big|_0^\infty} = 1 - J_0^2(v) - J_1^2(v) \qquad (16.9\text{-}25)$$

The above function is plotted in Fig. 16.18; one can deduce from the curve that about 84% of light is contained within the circle bounded by the first dark ring and about 91% of the light is contained in the circle bounded by the first two dark rings, etc.

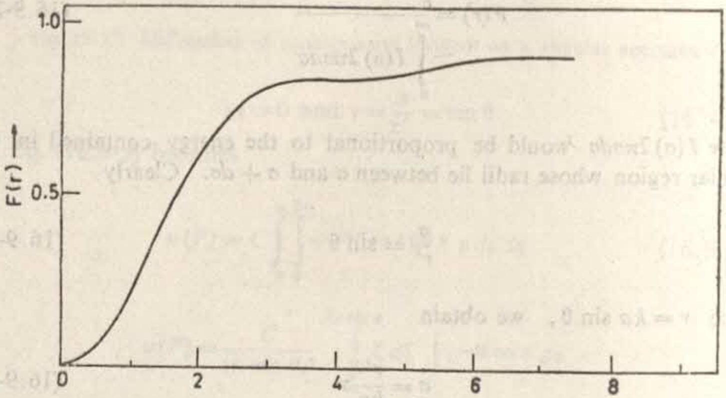


Fig. 16.18 The fractional energy contained in a circle of radius *r*.

PROBLEMS

1. Consider a circular aperture of radius 0.01 cm illuminated by a plane wave of wavelength 6×10^{-5} cm. Calculate the positions of the brightest and the darkest points on the axis.
2. What would happen if the circular aperture in Problem 1 is replaced by a circular disc of the same radius?
3. If a zone-plate has to have a principal focal length of 50 cm corresponding to $\lambda = 6 \times 10^{-5}$ cm, obtain an expression for the radii of different zones.
4. In a zone-plate, the second, fourth, sixth...zones are blackened; what would happen if instead the 1st, 3rd, 5th, etc. zones were blackened?
5. Consider a circular aperture of diameter 2 mm illuminated by a plane wave. The most intense point on the axis is at a distance of 200 cm from the aperture. Calculate the wavelength. (Ans. 5×10^{-5} cm).
6. Consider a straight edge being illuminated by a parallel beam of light with $\lambda = 6 \times 10^{-5}$ cm. Calculate the positions of the first few maxima and minima on a screen at a distance of 50 cm from the edge.
7. In the straight edge diffraction pattern, one observes that the most intense maximum occurs at a distance of 1 mm from the edge of the geometrical shadow.

Calculate the wavelength of light, if the distance between the screen and the straight edge is 300 cm. (Ans. $\simeq 4400$ Å).

8. In the straight edge diffraction pattern if the wavelength of the light used is 6000 Å and if the distance between the screen and the straight edge is 100 cm, calculate the distance between the most intense maximum and the next minimum.
9. Consider a plane wave falling normally on a narrow slit of width 0.2 cm. If the wavelength of light is 6×10^{-5} cm, study the diffraction pattern for various values of the distance between the slit and the screen. What should be this distance for the diffraction to be essentially Fraunhofer?
10. In Sec. 16.9 we obtained the diffraction pattern of a circular aperture of radius a. Obtain the diffraction pattern of an annular aperture bounded by circles of radii a_1 and $a_2 (> a_1)$. (Hint: The integration limits of ρ in Eq. (16.9–15) must be a_1 and a_2).
11. Consider a rectangular aperture of dimension 0.2 mm × 0.3 mm. Obtain the positions of the first few maxima and minima in the Fraunhofer diffraction pattern along directions parallel to the length and the breadth of the rectangle. Assume $\lambda = 5 \times 10^{-5}$ cm and that the diffraction pattern is produced at the focal plane of a lens of focal length 20 cm.
12. The Fraunhofer diffraction pattern of a circular aperture (of radius 0.5 mm) is observed on the focal plane of a convex lens of focal length 20 cm. Calculate the radii of the first and the second dark rings. Assume $\lambda = 5.5 \times 10^{-5}$ cm.
13. In the above problem, calculate the area of the patch (on the focal plane) which will contain 95% of the total energy. [Hint: Use Eq. (16.9–25).]
14. A plane wave is incident on a convex lens of focal length 100 cm. If the diameter of the lens is 6 cm, calculate the radius of the first Airy ring. Assume $\lambda = 6 \times 10^{-5}$ cm. If the diameter of the lens is increased to 20 cm, what would be the radius of the first Airy ring? Interpret the result physically.

REFERENCES AND SUGGESTED READING

BALL, C.J. (1971), *An Introduction to the Theory of Diffraction*, Pergamon Press, Oxford (England).

BORN, M. AND WOLF, E. (1975), *Principles of Optics*, Pergamon Press, Oxford (England).

BRADDICK, H.J.J. (1965), *Vibration, Waves and Diffraction*, McGraw-Hill Publishing Company, London (England).

BRAGG, L. (1968), X-Ray Crystallography, *Scientific American*, July Issue.

COWLEY, J.M. (1975), *Diffraction Physics*, North Holland Publishing Co., Amsterdam (Netherlands).

ELMORE, W.C. AND HEALD, M.A. (1969), *Physics of Waves*, McGraw-Hill Book Co., New York (USA).

FRANCON, M. (1966), *Diffraction, Coherence In Optics*, Pergamon Press, Oxford (England).

GARMIRE, E., CHIAO, R.V. AND TOWNES, C.H. (1966), Dynamics of Characteristics of the Self-trapping of Intense Light Beams, *Phys. Rev. Lett.*, **16**, 347.

GOODMAN, J.W. (1968), *Introduction to Fourier Optics*, McGraw-Hill Book Co., New York (USA).

JENKINS, F.A. AND WHITE H.E. (1976), *Fundamentals of Optics*, McGraw-Hill Book Co., New York (USA).

MEINERS, H.F. (1970), *Physics: Demonstration and Experiments*, (Vol. 2), The Ronald Press Co., New York (USA).

PANOFSKY, W.K.H. AND PHILIPS, M. (1962), *Classical Electricity and Magnetism*, Addison Wesley Publishing Co., Reading, Mass. (USA).

PHILLIPS, R.A. (1969), Spatial Filtering Experiments for Undergraduate Laboratories, *Amer. J. Phys.*, 37, 536.

SODHA, M.S. (1973), Theory of Nonlinear Refraction: Self Focussing of Laser Beams, *J. Phys. Edu.* (India), 1(2), 13.

SODHA, M.S., GHATAK, A.K. AND TRIPATHI, V.K. (1974), *Self-focussing of Laser Beams in Dielectrics, Plasmas and Semiconductors*, Tata McGraw-Hill, New Delhi, (India).

STONE, J.M. (1963), *Radiation and Optics*, McGraw-Hill Book Co. New York (USA)

WAGNER, W.G., HAUS, H.A. AND MARBURGER, J.M. (1968), Large Scale Self-trapping of Optical Beams in the Paraxial Ray Approximation, *Phys. Rev.*, 175, 256.

17 Holography

Within 15 years...viewers may gaze, in their own living room at an eight-foot-tall image of John Wayne (presumably still going strong) and even walk around him.

LIFE

17.1 Introduction

A photograph represents a two-dimensional recording of a three-dimensional scene. What is recorded is the intensity distribution that prevailed at the plane of the photograph when it was exposed. The light sensitive medium is sensitive only to the intensity variations and hence while recording a photograph, the phase distribution which prevailed at the plane of the photograph is lost. Since only the intensity pattern has been recorded, the three-dimensional character (e.g., parallax) of the object scene is lost. Thus one cannot change the perspective of the image in the photograph by viewing it from a different angle or one cannot refocus any unfocused part of the image in the photograph. Holography is a method evolved by Gabor in 1947, in which one not only records the amplitude but also the phase of the light wave; this is done by using interferometric techniques. Because of this the image produced by the technique of holography has a true three-dimensional form. Thus, as with the object, one can change one's position and view a different perspective of the image or one can focus at different distances. The capability to produce images as true as the object itself is what is responsible for the wide popularity gained by holography.

The basic technique in holography is the following: In the recording of the hologram, one superimposes on the object wave another wave called the reference wave and the photographic plate is made to record the resulting interference pattern (see Fig. 17.1). The reference wave is

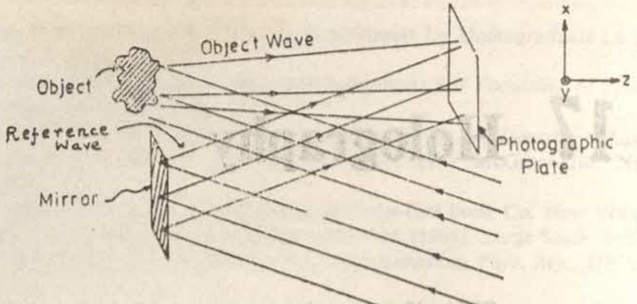


Fig. 17.1 Recording of a hologram.

usually a plane wave. This recorded interference pattern forms the hologram and (as will be shown) contains information not only about the amplitude but also about the phase of the object wave. Unlike a photograph, the hologram has little resemblance with the object; in fact, information about the object is coded into the hologram. To view the image, we again illuminate the hologram with another wave, called the reconstruction wave (which in most cases is identical to the reference wave used during the formation of the hologram); this process is termed reconstruction (see Fig. 17.2). The reconstruction process leads, in

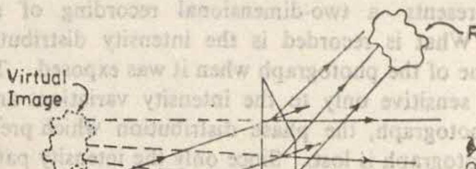
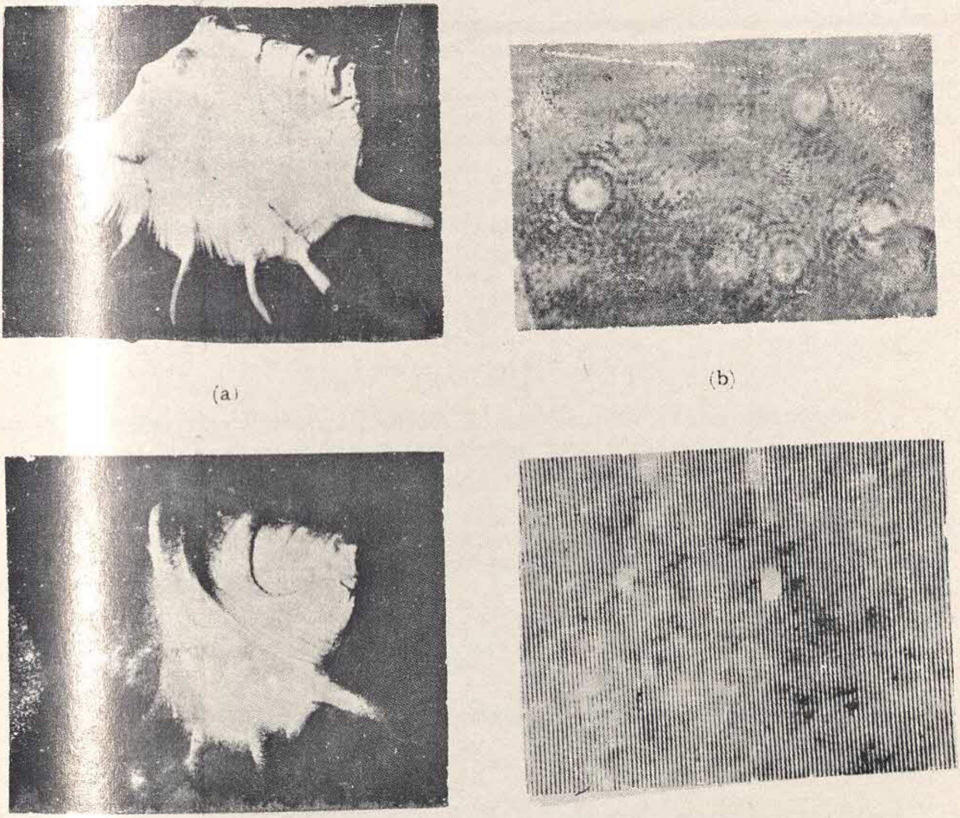


Fig. 17.2 Reconstruction process.

general, to a virtual and a real image of the object scene. The virtual image has all the characteristics of the object like parallax, etc. Thus one can move the position of the eye and look behind the objects or one can focus at different distances. The real image can be photographed without the aid of lenses just by placing a light sensitive medium at the position where the real image is formed. Plates 18 (a), (b) and (c) represent the object, its hologram and the reconstructed image respectively.



(a)

(b)

(c)

(d)

Plate 18 (a) An ordinary photograph of an object. (b) The hologram of the object produced by a method similar to the one as shown in Fig. 17.1. (c) The reconstructed image as seen by an observer. (d) A magnified view of a small portion of the hologram shown in (b). (Photographs courtesy Dr. R.S. Sirohi.)

17.2 Theory

If the object is a point scatterer, then the object wave would just be $\frac{A}{r}\cos(kr-\omega t+\varphi)$ where r represents the distance of the point of observation from the point scatterer and A represents a constant; $k=2\pi/\lambda$. Any general object can be thought of as being made up of a large number of points and the composite wave reflected by the object would be vectorial sum of these. The fundamental problem in holography is the recording of this object wave, in particular, the phase distribution associated with it.

Let us consider the recording process. Let

$$O(x,y)=a(x,y)\cos[\varphi(x,y)-\omega t] \qquad (17.2\text{-}1)$$

represent the object wave (which as mentioned earlier, is due to the superposition of waves from point scatterers on the object) in the plane of the photographic plate which is assumed to be $z=0$ (see Fig. 17.1). We consider a plane reference wave and assume, for simplicity, that it is propagating in the x-z-plane inclined at an angle θ with the z-direction (see Fig. 17.1). Thus, the field associated with this plane wave would be given by

$$\begin{aligned} r(x,y,z)&=A\cos[\mathbf{k}\cdot\mathbf{r}-\omega t]\\ &=A\cos(kx\sin\theta+kz\cos\theta-\omega t) \qquad (17.2\text{-}2)\end{aligned}$$

If $r(x,y)$ represents the field at the plane $z=0$ due to this reference wave, then one can see that

$$\begin{aligned} r(x,y)&=A\cos[kx\sin\theta-\omega t]\\ &=A\cos[2\pi\alpha x-\omega t] \qquad (17.2\text{-}3)\end{aligned}$$

where $\alpha=\sin\theta/\lambda$ is the spatial frequency (see Sec. 15.11). The above equation represents the field due to a plane wave inclined at an angle θ with the z-axis and as can be seen the phase varies linearly with x. Notice that there is no y-dependence because the plane wave has been assumed to have its propagation vector in the x-z-plane. Thus the total field at the photographic plate (which is coincident with the plane $z=0$) would be given by

$$\begin{aligned} u(x,y,t)&=a(x,y)\cos[\varphi(x,y)-\omega t]\\ &\quad+A\cos(2\pi\alpha x-\omega t) \qquad (17.2\text{-}4)\end{aligned}$$

The photographic plate responds only to the intensity which would be proportional to the time average of $[u(x,y,t)]^2$. Thus, the intensity pattern recorded by the photographic plate would be

$$\begin{aligned} I(x,y)&=\langle u^2(x,y,t)\rangle\\ &=\langle[a(x,y)\cos\{\varphi(x,y)-\omega t\}+A\cos(2\pi\alpha x-\omega t)]^2\rangle \qquad (17.2\text{-}5)\end{aligned}$$

where angular brackets denote time averaging (see Sec. 14.3). Thus

$$I(x, y) = a^2(x, y)\langle\cos^2\{\varphi(x, y) - \omega t\}\rangle + A^2\langle\cos^2(2\pi\alpha x - \omega t)\rangle + 2a(x, y)A\langle\cos\{\varphi(x, y) - \omega t\}\cos(2\pi\alpha x - \omega t)\rangle \qquad (17.2\text{-}6)$$

Since

$$\langle\cos^2[\varphi(x, y) - \omega t]\rangle = \tfrac{1}{2} = \langle\cos^2(2\pi\alpha x - \omega t)\rangle \qquad (17.2\text{-}7)$$

and

$$\begin{aligned}\langle\cos[\varphi(x, y) - \omega t]\cos(2\pi\alpha x - \omega t)\rangle &= \tfrac{1}{2}\langle\cos[\varphi(x, y) + 2\pi\alpha x - 2\omega t] + \cos[\varphi(x, y) - 2\pi\alpha x]\rangle \\ &= \tfrac{1}{2}\cos[\varphi(x, y) - 2\pi\alpha x], \qquad (17.2\text{-}8)\end{aligned}$$

Eq. (17.2-6) becomes

$$I(x, y) = \tfrac{1}{2}a^2(x, y) + \tfrac{1}{2}A^2 + Aa(x, y)\cos[\varphi(x, y) - 2\pi\alpha x] \qquad (17.2\text{-}9)$$

From the above relation it is obvious that the phase information of the object wave, which is contained in $\varphi(x, y)$, is recorded in the intensity pattern.

When the photographic plate which has recorded the above intensity pattern is developed one obtains a hologram (see Plates 18(b) and (d)). The transmittance of the hologram, i.e., the ratio of the transmitted field to the incident field, depends on $I(x, y)$. By a suitable developing process one can obtain a condition under which the amplitude transmittance would be linearly related to $I(x, y)$. Thus, in such a case if $R(x, y)$ represents the field of the reconstruction wave at the hologram plane then the transmitted field would be given by

$$\begin{aligned}v(x, y) &= KR(x, y)\,I(x, y) \\ &= K[\tfrac{1}{2}a^2(x, y) + \tfrac{1}{2}A^2]\,R(x, y) \\ &\qquad + KAa(x, y)\,R(x, y)\cos[\varphi(x, y) - 2\pi\alpha x] \qquad (17.2\text{-}10)\end{aligned}$$

where K is a constant. We consider the case when the reconstruction wave is identical to the reference wave $r(x, y)$ (see Fig. 17.2). In such a case we would obtain (omitting the constant K)

$$\begin{aligned}v(x, y) &= [\tfrac{1}{2}a^2(x, y) + \tfrac{1}{2}A^2]A\cos(2\pi\alpha x - \omega t) + A^2a(x, y) \\ &\qquad \times\cos(2\pi\alpha x - \omega t)\cos[\varphi(x, y) - 2\pi\alpha x] \\ &= [\tfrac{1}{2}a^2(x, y) + \tfrac{1}{2}A^2]A\cos(2\pi\alpha x - \omega t) + \tfrac{1}{2}A^2a(x\ y) \\ &\qquad \cos[\varphi(x, y) - \omega t] + \tfrac{1}{2}A^2a(x, y)\cos[4\pi\alpha x - \varphi(x, y) - \omega t] \\ &\qquad (17.2\text{-}11)\end{aligned}$$

The above equation gives the transmitted field in the plane $z = 0$. We consider each of the three terms separately. The first term is nothing but the reconstruction wave itself whose amplitude is modulated due to the presence of the term $a^2(x, y)$. This part of the total field is travelling in the direction of the reconstruction wave. The second term is identical (within a constant term) to the RHS of Eq. (17.2-1) and hence represents the original object wave; this gives rise to a virtual image. Thus the effect of viewing this wave is the same as viewing the object itself. The reconstructed object wave is travelling in the same direction as the original object wave.

To study the last term we first observe that in addition to the term $4\pi\alpha x$, the phase term $\varphi(x, y)$ carries a negative sign. The negative sign represents the fact that the wave has a curvature opposite to that of the object wave. Thus if the object wave is a diverging spherical wave then the last term represents a converging spherical wave. Thus in contrast to the second term, this wave forms a real image of the object which can be photographed by simply placing a film (see Fig. 17.2). To determine the effect of the term $4\pi\alpha x$, we consider the case when the object wave is also a plane wave travelling along the z-axis. For such a wave $\varphi(x, y) = 0$ and the last term would represent a plane wave propagating along a direction $\theta' = \sin^{-1}(2 \sin \theta)$. Thus the effect of the term $4\pi\alpha x$ is to rotate the direction of the wave. Hence the last term on the RHS of Eq. (17.2-11) represents the conjugate of the object wave propagating along a direction different from that of the reconstruction wave and the object wave, which forms a real image of the object. Since the waves represented by the three terms are propagating along different directions they separate after traversing a distance and enable the observer to view the virtual image without any disturbance.

A very interesting property possessed by holograms is that even if the hologram is broken up into different fragments, each separate fragment is capable of producing a complete virtual image of the object.* This property can be understood from the fact that for a diffusely reflecting object, each point of the object illuminates the complete hologram and consequently each point in the hologram receives waves from the complete object. But the resolution in the image decreases as the size of the fragment decreases. For non-diffusely reflecting objects or for transparencies, one makes use of an additional diffusing screen through which the object is illuminated.

Example 17.1 As an explicit example of the formation and reconstruction of a hologram, we consider the simple case when both the object wave and the reference wave are plane waves [see Fig. 17.3(a)]. (A plane object wave corresponds to a single object point lying far away from the hologram). (a) Show that for such a case, the hologram consists of a series of Young's interference fringes having an intensity distribution of the $\cos^2$ type. (b) If we reconstruct the hologram with another plane wave [see Fig. 17.3(b)], then show that the transmitted light consists of a zero-order plane wave and two first-order plane waves; the two first order waves correspond to the primary and conjugate waves.

---

*This property of a hologram exists only when the object is a diffuse scatterer such that the wave from each scattering point of the object reaches all parts of the hologram plate. There are cases where this does not hold good: for example, when a hologram of a transparency is to be recorded.

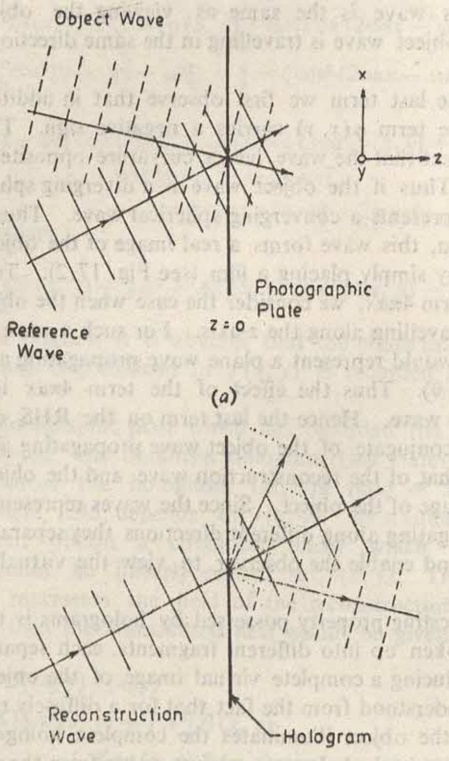


Fig. 17.3 (a) Formation of a hologram, when both the object wave and the reference waves are plane waves. (b) Reconstruction of the hologram with another plane wave.

Solution.

(a) Consider a plane wave with its propagation vector lying in the *x*-*z*-plane and making angle θ_1 with the *z*-axis. For such a wave, the field is of the form

$$A_1 \cos [kx \sin \theta_1 + kz \cos \theta_1 - \omega t]$$

If the photographic film is assumed to coincide with the plane $z = 0$, then the field distribution on this plane would be given by

$$A_1 \cos [kx \sin \theta_1 - \omega t]$$

Similarly the field (on the plane of the film) due to a plane wave making an angle θ_2 with the *z*-axis, will be

$$A_2 \cos [kx \sin \theta_2 - \omega t]$$

The resultant intensity distribution would be proportional to

$$\begin{aligned}\langle [A_1 \cos \{kx \sin \theta_1 - \omega t\} + A_2 \cos \{kx \sin \theta_2 - \omega t\}]^2 \rangle &= \tfrac{1}{2} A_1^2 + \tfrac{1}{2} A_2^2 + A_1 A_2 \cos [kx (\sin \theta_1 - \sin \theta_2)] \\ &= \tfrac{1}{2} (A_1 - A_2)^2 + 2 A_1 A_2 \cos^2 \left[\frac{kx}{2} (\sin \theta_1 - \sin \theta_2)\right]\end{aligned}$$

For $A_1 = A_2$, the above expression simplifies to

$$2 A^2 \cos^2 \left[\frac{kx}{2} (\sin \theta_1 - \sin \theta_2)\right]$$

showing that the intensity remains constant along lines parallel to the y-axis with fringe spacing depending on the values θ_1 and θ_2. Further, the intensity distribution is of the $\cos^2$ type (cf. Fig. 11.8 and Plate 6).

(b) Before we calculate the transmitted field of the hologram, we first consider a single slit of width b being illuminated by a plane wave. Consider an element ds at a distance s from the centre of the slit. Then the amplitude at an angle θ due to this element would be proportional to $\sin [kr + ks \sin \theta - \omega t]\, ds$; here $k = 2\pi/\lambda$. Thus the total field in the direction θ would be given by

$$E = A \int_{-b/2}^{b/2} \sin [kr + ks \sin \theta - \omega t]\, ds \tag{17.2-12}$$

where A is a constant. The above integral can also be written as

$$\begin{aligned}E &= A \int_{-b/2}^{b/2} [\sin (kr - \omega t) \cos (ks \sin \theta) - \cos (kr - \omega t) \sin (ks \sin \theta)]\, ds \\ &= 2A \sin (kr - \omega t) \frac{\sin\left(\dfrac{kb}{2} \sin \theta\right)}{k \sin \theta} \qquad (17.2\text{-}13) \\ &= Ab \sin (kr - \omega t) \left\{\sin\left(\frac{kb}{2} \sin \theta\right) \Big/ \frac{kb}{2} \sin \theta\right\}\end{aligned}$$

where the second integral is zero because the integrand is an odd function of s. The amplitude is of the form $\left\{\sin\left(\frac{kb}{2} \sin \theta\right) \Big/ \left(\frac{kb}{2} \sin \theta\right)\right\}$, the same as obtained in Sec. 15.2.

In the present case, the hologram has a $\cos^2 \alpha s$ type of variation in transmittance and hence the transmitted field will be of the form

$$E = A \int_{-b/2}^{b/2} \cos^2 \alpha s \sin [kr + ks(\sin \theta - \sin \theta_i) - \omega t]\, ds \tag{17.2-14}$$

where θ_i represents the angle of incidence of the illuminating plane wave. Thus

$$E = \tfrac{1}{2} A \int_{-b/2}^{b/2} (1 + \cos 2\alpha s)\, [\sin (kr - \omega t) \cos \{ks(\sin \theta - \sin \theta_i)\} + \cos (kr - \omega t) \sin \{ks(\sin \theta - \sin \theta_i)\}]\, ds$$

$$= \tfrac{1}{2} A \sin (kr - \omega t) \left[\int_{-b/2}^{b/2} \cos \{ks(\sin \theta - \sin \theta_i)\}\, ds + \tfrac{1}{2} \int_{-b/2}^{b/2} \cos \{ks(\sin \theta - \sin \theta_i + 2\alpha)\}\, ds + \tfrac{1}{2} \int_{-b/2}^{b/2} \cos \{ks(\sin \theta - \sin \theta_i - 2\alpha)\}\, ds \right] \tag{17.2-15}$$

The above integrations can easily be carried out. Thus, for example,

$$\int_{-b/2}^{b/2} \cos \{ks(\sin \theta - \sin \theta_i + 2\alpha)\}\, ds = \frac{\sin \left[b \dfrac{k}{2} \left(\sin \theta - \sin \theta_i + 2\alpha \right) \right]}{\dfrac{k}{2} (\sin \theta - \sin \theta_i + 2\alpha)} \tag{17.2-16}$$

which becomes more and more sharply peaked around $\sin \theta = \sin \theta_i - 2\alpha$ as $b \rightarrow \infty$, i.e. as the size of the hologram becomes larger. Thus the three integrals in Eq. (17.2-15) in the limit of a large value of b give rise to three plane waves propagating along $\sin \theta = \sin \theta_i$, $\sin \theta = \sin \theta_i - 2\alpha$ and $\sin \theta = \sin \theta_i + 2\alpha$, which represent the zero-order and the two first order waves.

Example 17.2 Consider the formation of a hologram with a point object and a plane reference wave. Choose the z-axis to be along the normal from the point source to the plane of the photograph, assumed to be coincident with the plane $z = 0$. For simplicity assume the reference wave to fall normally on the photographic plate. Obtain the interference pattern recorded by the hologram.

Solution. Let the point source be situated at a distance d from the photographic plate. The field at any point $P(x, y, 0)$ on the photographic plate, due to waves emanating from the point object would be given by

$$O(x, y, z=0, t) = \frac{A}{r} \cos(kr - \omega t) \qquad (17.2\text{-}17)$$

where $r = (x^2 + y^2 + d^2)^{1/2}$ and A represents a constant. A plane wave travelling along a direction parallel to the z-axis would be given by

$$R(x, y, z, t) = B \cos(kz - \omega t) \qquad (17.2\text{-}18)$$

Hence, the field due to the reference wave at the plane of the photographic plate ($z = 0$) would be

$$R(x, y, z=0, t) = B \cos \omega t \qquad (17.2\text{-}19)$$

Thus, the total field at the plane of the photographic plate would be

$$\begin{aligned} T(x, y, t) &= O(x, y, z=0, t) + R(x, y, z=0, t) \\ &= \frac{A}{r} \cos(kr - \omega t) + B \cos \omega t \qquad (17.2\text{-}20) \end{aligned}$$

The recorded intensity pattern would be

$$\begin{aligned} I(x, y) &= \langle |T(x, y, t)|^2 \rangle \\ &= \left\langle \left| \frac{A}{r} \cos(kr - \omega t) + B \cos \omega t \right|^2 \right\rangle \end{aligned}$$

where, as before, angular brackets denote time averaging. Carrying out the above time averaging, we get

$$I(x, y) = \frac{A^2}{2r^2} + \frac{B^2}{2} + \frac{AB}{r} \cos kr \qquad (17.2\text{-}21)$$

If we assume that $d \gg x, y$ (which is valid in most practical cases), we can write

$$r = (x^2 + y^2 + d^2)^{1/2} \simeq d + \frac{x^2 + y^2}{2d} \qquad (17.2\text{-}22)$$

Thus

$$I(x, y) = \frac{A^2}{2r^2} + \frac{B^2}{2} + \frac{AB}{r} \cos\left[kd + \frac{k}{2d}\left(x^2 + y^2\right)\right] \qquad (17.2\text{-}23)$$

The resultant fringe pattern is circular and centred at the origin. Bright fringes occur at radii given by

$$r_m = \left[\frac{2d}{k} 2m\pi\right]^{1/2} = \sqrt{2m}\sqrt{\lambda d} \qquad (17.2\text{-}24)$$

where we have omitted a constant phase factor kd. Thus, the radii of the bright fringes are proportional to the square root of even integers. Similarly, the radii of dark rings would be proportional to the square root of odd integers. The hologram thus formed is essentially a zone plate with the transmittance varying sinusoidally in contrast to the Fresnel zone plate (see Sec. 16.3).

17.3 Requirements

Since holography is essentially an interference phenomenon, certain coherence requirements have to be met with. In Chapter 14 we had introduced the notion of coherence length. Thus, if stable interference fringes are to be formed (so that they are recordable), the maximum path difference between the object wave and the reference wave should not exceed the coherence length. Further, the spatial coherence is important so that the waves scattered from different regions of the object could interfere with the reference beam.

During reconstruction, the reconstructed image coordinates depend both on the wavelength and the position of the reconstructing source. Hence if the resolution in the reconstructed image has to be good, the source must not be broad and must be emitting a narrow band of wavelengths. It may be worthwhile mentioning here that the reconstruction process has associated with it aberrations similar to that in the images formed by lenses. If the reconstruction source is of the same wavelength and is situated at the same relative position with respect to the hologram as the reference source, then the reconstructed image does not suffer from any aberrations.

Another critical requirement in making holograms is stability of the recording arrangement. Thus, the film, the object and any mirrors used in producing the reference beam must be motionless with respect to one another during the exposure. One more requirement which is not so obvious (but is a necessity) is the resolution of the film. Two plane waves making angles $+\theta$ and $-\theta$ with the axis, produce an interference pattern with spacing $d=\frac{2\sin\theta}{\lambda}$. Assuming $\theta=15°$ and $\lambda=6328$ Å (He-Ne laser), one obtains $d=1.222\times10^{-3}$ mm; thus the spatial frequency is 818 lines/mm. Thus the photographic plate should be able to record fringes as close as 0.1222×10^{-4} mm apart. This requires special kinds of material which tend to be exceedingly slow thus taking the stability requirements even further. Some of the holographic materials are 649 Kodak or 10*E* or 8*E* 75 Agfa-Gaevert films and plates.

17.4 Some Applications

The principle of holography finds applications in many diverse fields. The ability to record information about the depth finds application in studying transient microscopic events. Thus, if one has to study some transient phenomenon which occurs in a certain volume, then using ordinary microscopic techniques it becomes difficult to first locate the position and make observation. If a hologram is recorded of the scene, then the event gets frozen into the hologram and hence one can focus through the depth of the reconstructed image and study the phenomenon at leisure.

One of the most promising applications of holography lies in the field of interferometry. The ability of the holographic process to release the object wave when reconstructed with a reconstruction wave allows us to perform interference between different waves which exist at different times. Thus, in the technique called double exposure holographic interferometry, the photographic plate is first partially exposed to the object wave and the reference wave. Then, the object is stressed and the photographic plate is again exposed along with the same reference wave. The photographic plate after development forms the hologram. When this hologram is illuminated with a reconstruction wave, then two object waves emerge from the hologram; one of them corresponds to the unstressed object and the other to the stressed object. Since the object waves themselves have been reconstructed, they interfere and produce interference fringes. These interference fringes are characteristic of the strain suffered by the body. A quantitative study of the fringe pattern produced in the body gives the distribution of strain in the object.

To understand the formation of the fringe pattern, we assume that the deformation of the object has been such as to alter only the phase distribution. Thus, if

$$O(x, y, t) = A(x, y) \cos[\varphi(x, y) - \omega t] \qquad (17.4\text{-}1)$$

represents the object wave (in the hologram plane) when the object is unstressed [see Fig. 17.4(a)] and if $O'(x, y, t)$ represents the object wave when the object is stressed [see Fig. 17.4(b)] then we may write

$$O'(x, y, t) = A(x, y) \cos[\varphi'(x, y) - \omega t] \qquad (17.4\text{-}2)$$

where the phase distribution has been assumed to change from $\varphi(x, y)$ to $\varphi'(x, y)$. On reconstruction, each of the above two object waves emerge from the hologram and what would be observed will be the intensity pattern due to interference of the two waves which would be given by*

$$\begin{aligned} I(x, y) &= < [A(x, y) \cos[\varphi(x, y) - \omega t] + A(x, y) \cos\{\varphi'(x, y) - \omega t\}]^2 > \\ &= A^2(x, y) + A^2(x, y) \cos[\varphi'(x, y) - \varphi(x, y)] \end{aligned} \qquad (17.4\text{-}3)$$

Thus, whenever,

$$\varphi'(x, y) - \varphi(x, y) = 2m\pi; \quad m = 0, 1, 2 \ldots \qquad (17.4\text{-}4)$$

the two waves would interfere constructively and whenever,

---

*The reconstruction process produces other wave components also but as was observed earlier, these components travel along different directions. Here we are concerned only with the object waves.

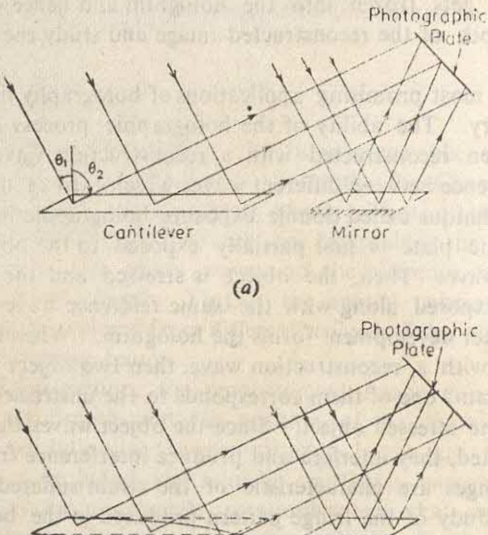


Fig. 17.4 (a) Recording of the unstressed object wave. (b) Recording of the stressed object wave on the same emulsion to produce the doubly exposed hologram.

$$\varphi'(x, y) - \varphi(x, y) = (2m + 1)\frac{\pi}{2}; \quad m = 0, 1, 2 \ldots . \qquad (17.4\text{-}5)$$

the two waves interfere destructively. Thus, depending on $[\varphi'(x, y) - \varphi(x, y)]$, one obtains, on reconstruction, the object superimposed with bright and dark fringes (see Plate 19).

We will consider here a simple application of the above technique in the determination of the Young's modulus of a material. If we have a bar fixed at one end and loaded at the other and if it results in a displacement δ of the end of the bar, then one can show that [see, e.g. Worsnop and Flint (1951), p. 75]

$$\delta = \frac{WL^3}{3YI}, \qquad (17.4\text{-}6)$$

where W is the load, L is the length of the bar, I is the moment of inertia of cross-section which for a rectangular bar of width a and thickness b, is given by $I = ab^3/12$; Y represents the Young's modulus of the material of the rod. Thus if we could determine δ for a given load, then Y can be determined from Eq. (17.4-6).

We will first determine an expression for $(\varphi' - \varphi)$. In Fig. 17.4 we

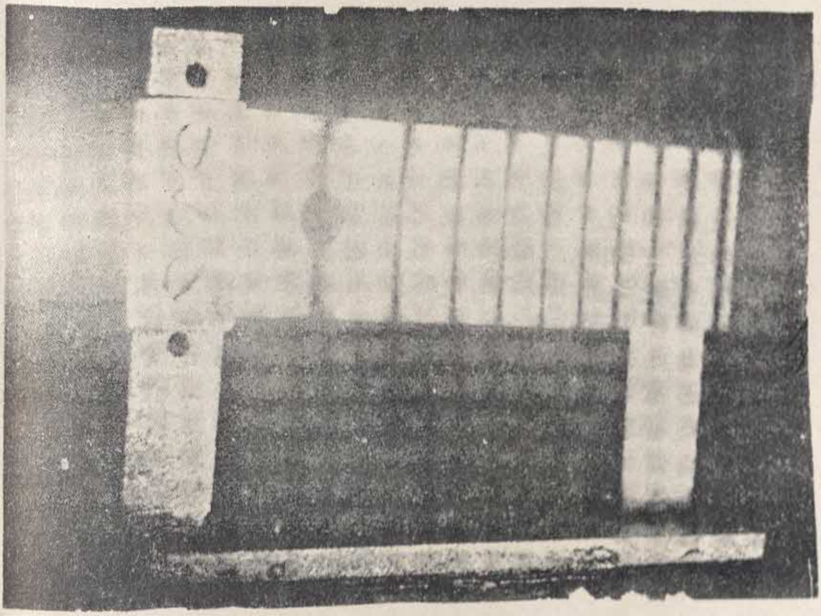


Plate 19 Interference fringes produced in the measurement of Young's modulus using double exposure interferometry. (Photograph courtesy Dr. R. S. Sirohi.)

have shown the undisplaced and displaced positions of the cantilever illuminated by a laser light along a direction making an angle θ_1 with the z-axis. We observe the cantilever along a direction making an angle θ_2 with the z-axis. The phase change when the cantilever undergoes a displacement δ as shown in Fig. 17.4(b) would be

$$\begin{aligned} \varphi' - \varphi &= \frac{2\pi}{\lambda}(\delta \cos\theta_1 + \delta \cos\theta_2) \\ &= \frac{2\pi}{\lambda}\delta(\cos\theta_1 + \cos\theta_2) \qquad (17.4\text{-}7) \end{aligned}$$

If there are N fringes over the length L of the cantilever, then since a phase difference of 2π corresponds to one fringe [see Eq. (17.4-4)] we can write

$$\frac{2\pi}{\lambda}\delta(\cos\theta_1 + \cos\theta_2) = N \cdot 2\pi$$

or

$$\delta = \frac{N\lambda}{(\cos\theta_1 + \cos\theta_2)}$$

Thus by measuring N, θ_1 and θ_2 and knowing λ, δ can be determined.

Plate 19 shows the reconstruction of a double exposed hologram of an aluminum strip of width 4 cm, thickness 0.2 cm and of length 12 cm. From the number of fringes formed, one can calculate the Young's modulus (see Problem 3).

PROBLEMS

1. What is the difference between the intensity distribution in a Fresnel zone-plate (see Sec. 16.3) and the one given by Eq. (17.2-23)?
2. Consider the reconstruction of the hologram as formed in the configuration of Example 17.2, by a plane wave travelling along a direction parallel to the z-axis. Show the formation of a virtual and a real image.
3. Plate 19 corresponds to the reconstruction of a doubly exposed hologram, the objects corresponding to the unstrained and strained positions of an aluminum bar of width 4 cm, thickness 0.2 cm and length 12 cm. If the strained position corresponds to a load of 1 gm force applied at the end of the bar, calculate the Young's modulus of aluminium. Assume $\theta_1 \approx \theta_2 \approx 0$ (Hint: N represents the number of fringes produced over the length of the cantilever).
(Ans. 0.7×10^{11} N/m^2)

REFERENCES AND SUGGESTED READING

Brown, J.C. and Harte, J.A. (1969). Holography in the undergraduate optics course, *Amer. J. Phys.*, 37, 441.

Caulfield, H.J. and Lu, S. (1970), *The Applications of Holography*, John Wiley and Sons, New York (USA).

COLLIER, R.J., BURCKHARDT, C.B. AND LIN, L.H. (1971), *Optical Holography*, Academic Press, New York (USA).

GIVENS, M.P. (1967), Introduction to holography, *Amer. J. Phys.*, **35**, 1056.

LEITH, E.N. AND UPATNIEKS, J. (1965), Photography by laser, *Scient. Amer.*, **212**, (June) p. 24.

METHERNAL, A.F. (1969), Acoustical holography, *Scient. Amer.* **221**, (October) p. 36.

PENNINGTON, K.S. (1968), Advances in holography, *Scient. Amer.*, **218** (February) p. 40.

SMITH, H.M. (1975), *Principles of Holography*, Wiley Interscience, New York (USA).

VENKATESHWARULU, D. (1970), Holography, theory and applications, *J. Scient. Ind. Res.*, **29**, (November).

WORSNOP, B.L. AND FLINT, H.T. (1951), *Advanced Practical Physics for Students*, Asia Publishing House, Bombay (India).

PART V

Electromagnetic Character of Light

18 Polarization and Double Refraction

> *...As to the other emanation which should produce the irregular refraction, I wished to try what Elliptical waves, or rather spheroidal waves, would do; and these I supposed would spread differently both in the ethereal matter diffused throughout the crystal and in the particles of which it is composed...*
>
> CHRISTIAAN HUYGENS

18.1 Introduction

If we move one end of a string up and down then a transverse wave is generated [see Fig. 18.1 (a)]. Each point of the string executes a sinusoidal oscillation in a straight line (along the x-axis) and the wave is, therefore, known as a linearly polarized wave. It is also known as a plane polarized wave because the string is always confined to the x-z plane. The displacement for such a wave can be written in the form

$$x(z, t) = a \cos (kz - \omega t + \varphi_1) \qquad (18.1\text{-}1)$$

where a represents the amplitude of the wave and φ_1 is the phase constant to be determined from the boundary condition; the y-coordinate of the displacement is always zero. At any instant the displacement will be a cosine curve as shown in Fig. 18.1 (a). Further, an arbitrary point $z = z_0$ will execute a simple harmonic motion with amplitude a. The string can also be made to vibrate in the y-z-plane [see Fig. 18.1 (b)] for which the displacement would be given by

$$y(z, t) = a \cos (kz - \omega t + \varphi_2) \qquad (18.1\text{-}2)$$

and

$$x(z, t) = 0$$

In general, the string can be made to vibrate in any plane containing the z-axis.

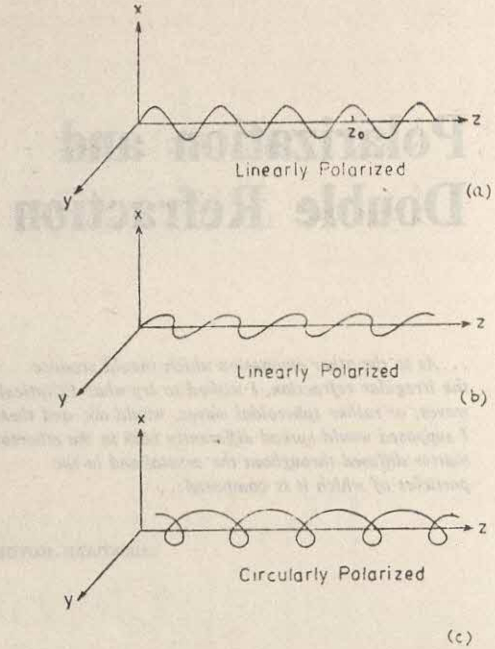


Fig. 18.1 (a) A linearly polarized wave on a string with the displacement confined to the x-z-plane; (b) A linearly polarized wave on a string with the displacement confined to the y-z plane; (c) A circularly polarized wave on a string.

If one rotates the end of the string on the circumference of a circle then each point of the string will move in a circular path as shown in Fig. 18.1 (c); such a wave is known as a circularly polarized wave and the corresponding displacement would be given by

$$x(z, t) = \frac{a}{\sqrt{2}} \cos(kz - \omega t + \varphi)$$

$$y(z, t) = \frac{a}{\sqrt{2}} \cos(kz - \omega t + \varphi) \qquad (18.1\text{-}3)$$

We next consider a long narrow slit S_1 placed in the path of the string as shown in Fig. 18.2 (a). If the length of the slit is along the direction of the displacement then the entire amplitude will be transmitted as shown in Fig. 18.2 (a). On the other hand, if the slit is at right angles to the direction of the displacement then almost nothing will be transmitted to the other side of the slit [see Fig. 18.2 (b)]. This is because of the fact that the slit allows only the component of the displacement, which is along the length of the slit, to pass through [see Fig. 18.2 (c)]. We must mention here that if a longitudinal wave was propagating through the string then the amplitude of the transmitted wave would have been the same for all orientations of the slit. Thus, the change in the amplitude of the transmitted wave with the orientation of the slit is due to the transverse character of the wave. Indeed, an experiment which is, in principle, very similar to the experiment discussed above proves the transverse character of light waves. However, before we discuss the experiment with light waves we must define an unpolarized wave. We once again consider transverse waves generated at one end of a string. If the plane of vibration is changed in a random manner in very short

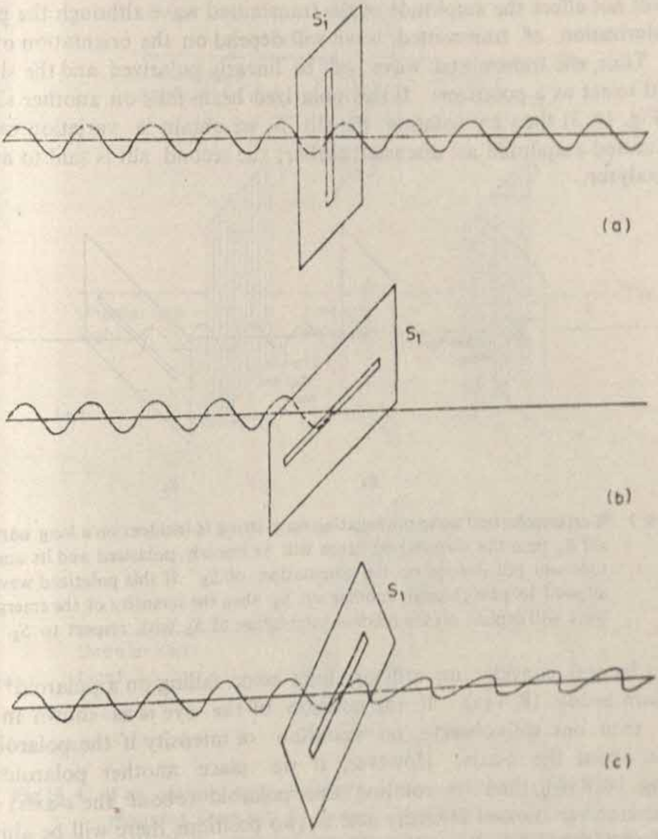


Fig. 18.2 If a linearly polarized transverse wave (propagating on a string) is incident on a long narrow slit, then the slit will allow only the component of the displacement, which is along the length of the slit, to pass through.

intervals of time, then such a wave is known as an unpolarized wave.* If an unpolarized wave falls on a slit then the displacement associated with the transmitted wave will be along the length of the slit and a rotation of the slit will not effect the amplitude of the transmitted wave although the plane of polarization of transmitted wave will depend on the orientation of the slit. Thus, the transmitted wave will be linearly polarized and the slit S_1 is said to act as a polarizer. If this polarized beam falls on another slit S_2 (see Fig. 18.3) then by rotating the slit S_2 we obtain a variation of the transmitted amplitude as discussed earlier; the second slit is said to act as an analyzer.

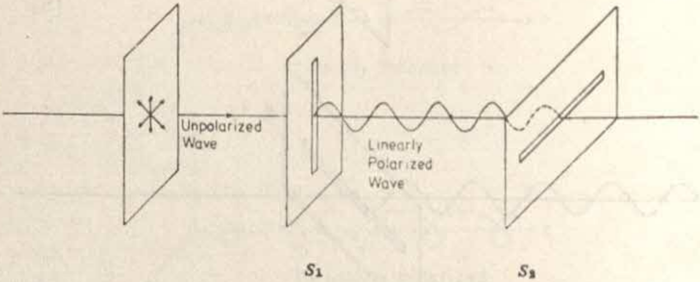


Fig 18.3 If an unpolarized wave propagating on a string is incident on a long narrow slit S_1 then the transmitted beam will be linearly polarized and its amplitude will not depend on the orientation of S_1. If this polarized wave is allowed to pass through another slit S_2, then the intensity of the emerging wave will depend on the relative orientation of S_2 with respect to S_1.

Let us next consider an ordinary light beam falling on a polaroid** P_1 as shown in Fig. 18.4 (a). If the position of the eye is as shown in the figure, then one will observe no variation of intensity if the polaroid is rotated about the z-axis. However, if we place another polaroid P_2 [see Fig. 18.4 (b)], then by rotating this polaroid (about the z-axis) one will observe variation of intensity and at two positions there will be almost complete darkness [see Fig. 18.4 (c)]. A similar phenomenon will also be observed if instead of rotating this polaroid P_2 we rotate P_1. On the basis of our earlier discussions, this phenomenon proves the transverse character of light; i.e. the displacement associated with a light wave is at right angles to the direction of propagation of the wave. The polaroid P_1 acts as a polarizer and the transmitted beam is linearly polarized. The second polaroid acts as an analyzer.

*By a short interval, we imply times which are short compared to the detection time; however, for the wave to be characterized with a certain frequency, ν this time has to be much greater than $1/\nu$, so that in the short interval it executes a large number of oscillations (see also Sec. 14-1).

**A polaroid is a plastic-like material used for producing polarized light; it will be discussed in detail in the next section. (Instead of a polaroid, we could also use a tourmaline crystal).

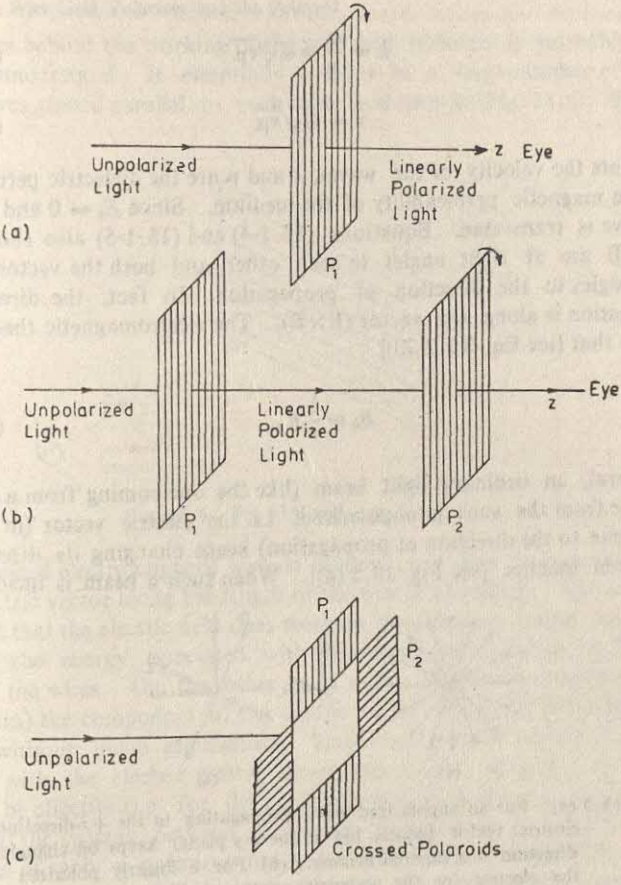


Fig 18.4 If an ordinary light beam is allowed to fall on a polaroid, then the emerging beam will be linearly polarized and if we place another polaroid P_2, then the intensity of the transmitted light will depend on the relative orientation of P_2 with respect to P_1.

The transverse character of light waves was known in the early years of the nineteenth century; however, the nature of the displacement associated with a light wave was known only after Maxwell had put forward his famous electromagnetic theory. We will discuss the basic electromagnetic theory in the next chapter where we will show that associated with a plane electromagnetic wave there is an electric field **E** and a magnetic field **B** which are at right angles to each other. For a linearly polarized wave propagating in the z-direction the electric and magnetic fields can be written in the form

$$E_x = E_0 \cos(kz - \omega t),\ E_y = 0,\ E_z = 0 \qquad (18.1\text{-}4)$$

and

$$B_x = 0,\ B_y = B_0 \cos(kz - \omega t),\ B_z = 0 \qquad (18.1\text{-}5)$$

where

$$k = \frac{\omega}{v} = \omega\sqrt{\epsilon\mu} \tag{18.1-6}$$

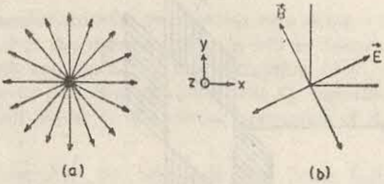
and

$$v = 1/\sqrt{\epsilon\mu} \tag{18.1-7}$$

represents the velocity of the waves, ϵ and μ are the dielectric permittivity and the magnetic permeability of the medium. Since $E_z = 0$ and $B_z = 0$, the wave is transverse. Equations (18.1-4) and (18.1-5) also show that **E** and **B** are at right angles to each other and both the vectors are at right angles to the direction of propagation. In fact, the direction of propagation is along the vector (**E**×**B**). The electromagnetic theory also tells us that [see Eq. (19.2-20)]

$$B_0 = \frac{1}{v} E_0 \tag{18.1-8}$$

In general, an ordinary light beam (like the one coming from a sodium lamp or from the sun) is unpolarized, i.e. the electric vector (in a plane transverse to the direction of propagation) keeps changing its direction in a random manner [see Fig. 18.5 (a)]. When such a beam is incident on a polaroid the emergent light is linearly polarized with its electric vector oscillating in a particular direction as shown in Fig. 18.5 (b). The direction of the electric vector of the emergent beam will depend on the orientation of the polaroid. As will be shown in Sec. 18.2 the component of **E** along a particular direction gets absorbed by the polaroid and the component at right angles to it passes through. The direction of the electric vector of the emergent wave is usually called the pass axis of the polaroid.



(a) (b)

Fig. 18.5 (a) For an unpolarized wave propagating in the +z-direction, the electric vector (which lies in the x-y plane) keeps on changing its direction in a random manner. (b) For a linearly polarized wave, the electric (or the magnetic) vector oscillates along a particular direction.

18.2 Production of Polarized Light

In this section we will discuss various methods for the production of linearly polarized light waves.

18.2.1 The Wire Grid Polarizer and the Polaroid

The physics behind the working of the wire grid polarizer is probably the easiest to understand. It essentially consists of a large number of thin copper wires placed parallel to each other as shown in Fig. 18.6. When an unpolarized electromagnetic wave is incident on it then the component of the electric vector along the length of the wire is absorbed. This is due to the fact that the electric field does work on the electrons inside the thin wires and the energy associated with the electric field is lost in the Joule heating of the wires. On the other hand, (since the wires are assumed to be very thin) the component of the electric vector along the x-axis passes through without much attenuation. Thus the emergent wave is linearly polarized with the electric vector along the x-axis. However, for the system to be effective (i.e. for the E_y component to be almost completely attenuated) the spacing between the wires should be $\lesssim \lambda$. Clearly, the fabrication of such a polarizer for a 3 cm microwave is relatively easy because the spacing has to be $\lesssim$ 3 cm. On the other hand, since the light waves are associated with a very small wavelength ($\sim 5\times10^{-5}$ cm), the fabrication of a polarizer in which the wires are placed at distances $\lesssim 5\times10^{-5}$ cm is extremely difficult. Nevertheless, Bird and Parrish did succeed in putting about 30,000 wires in about one inch.*

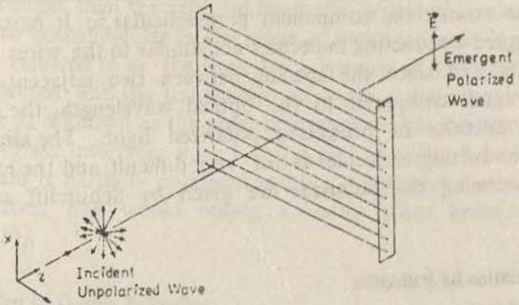


Fig. 18.6 The wire-grid polarizer.

As already pointed out, it is extremely difficult to fabricate a wire grid polarizer which would be effective for visible light. However, instead of long thin wires, one may employ long chain polymer molecules that contain atoms (like iodine) which provide high conductivity along the length of the chain. These long chain molecules are aligned so that they are almost parallel to each other. Because of the high conductivity

---

*For further details see Shurcliff and Ballard (1964). The details of the procedure for making this wire grating are also discussed in this book. The original work of Bird and Parrish was published in the *Journal of the Optical Society of America*, **50**, 886 (1950).

provided by the iodine atoms, the electric field parallel to the molecules gets absorbed. A sheet containing such long chain polymer molecules (which are aligned parallel to each other) is known as a polaroid. When a light beam is incident on such a polaroid, the molecules (aligned parallel to each other), absorb the component of the electric field which is parallel to the direction of the alignment because of the high conductivity provided by the iodine atoms; the component perpendicular to it passes through. Thus the aligned conducting molecules act similar to the wires in the wire grid polarizer and since the spacing between two adjacent long chain molecules is small compared to the optical wavelength, the polaroid is usually very effective in producing polarized light. The aligning of the long chain conducting molecules is not very difficult and the experimental details of producing the polarizer are given by Schurcliff and Ballard (1964).

18.2.2 Polarization by Reflection

Let us consider the incidence of a plane wave on a dielectric. We assume that the electric vector associated with the incident wave lies in the plane

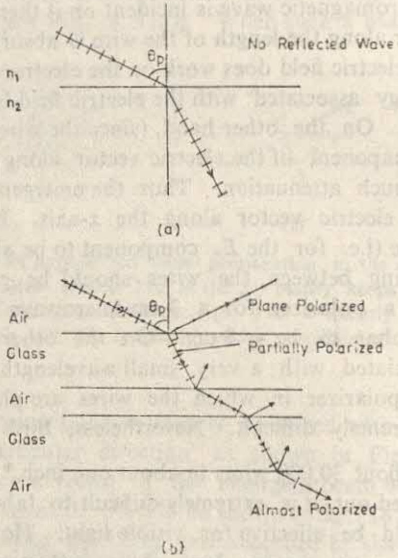


Fig. 18.7 (a) If a linearly polarized wave (with its E in the plane of incidence) is incident on the interface of two dielectrics with the angle of incidence equal to θ_p ($= \tan^{-1} n_2/n_1$) then the reflection coefficient is zero. (b) Thus, if an unpolarized beam is incident at this angle, the reflected beam is plane polarized whose electric vector is perpendicular to the plane of incidence. The transmitted beam is partially polarized and if this beam is made to undergo several reflections, then the emergent beam is almost plane polarized with its electric vector in the plane of incidence.

of incidence as shown in Fig. 18.7(a). It will be shown in Sec. 20.2 that if the angle of incidence (θ) is such that

$$\theta = \theta_p = \tan^{-1}\left(\frac{n_2}{n_1}\right) \tag{18.2-1}$$

then the reflection coefficient is zero. Thus, if an unpolarized beam is incident at this angle, then the reflected beam will be linearly polarized with its electric vector perpendicular to the plane of incidence [see Fig. 18.7(b)]. Eq. (18.2-1) is referred to as the Brewster's law and at this angle of incidence, the reflected and the transmitted rays are at right angles to each other; the angle θ_p is known as the polarizing angle or the Brewster angle.

For the air-glass interface, $n_1 = 1$ and $n_2 \simeq 1.5$ giving $\theta_p \simeq 57°$. The transmitted beam is partially polarized and if one uses a large number of reflecting surfaces, one would obtain a nearly plane polarized beam [see Fig. 18.7(b)].

18.2.3 Polarization by Double Refraction

In Secs. 18.5 and 18.9 we will discuss the phenomenon of double refraction and will show that when an unpolarized beam enters an anisotropic crystal, it splits up into two beams each of them being characterized by a certain state of polarization. If by some method, we could eliminate one of the beams then we would obtain a linearly polarized beam..

A simple method for eliminating one of the beams is through selective absorption.* A crystal like tourmaline has different coefficients of

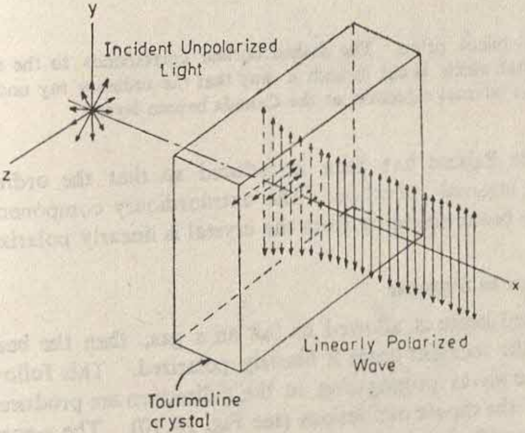


Fig. 18.8 When an unpolarized beam enters a dichroic crystal like tourmaline, it splits up into two linearly polarized components. One of the components gets absorbed quickly and the other component passes through without much attenuation. [Adapted from Alonso and Finn, 1970; used with permission.]

---

*This property of selective absorption is known as dichroism.

absortion for the two linearly polarized beams into which the incident beam splits up. Consequently, one of the beams gets absorbed quickly and the other component passes through without much attenuation. Thus, if an unpolarized beam is passed through a tourmaline crystal, the emergent beam will be linearly polarized (see Fig. 18.8).

Another method for eliminating one of the polarized beams is through total internal reflection. We will show in Sec. 18.5 and 18.9 that the two beams have different ray velocities and as such the corresponding refrative indices will be different. If one can sandwich a layer of a material whose refractive index lies between the two, then for one of the beams, the incidence will be at a rarer medium and for the other it will be at a denser medium. This principle is used in a Nicol prism which consists of a calcite crystal cut in such a way that for the beam, for which sandwiched material is a rarer medium, the angle of incidence is greater than the critical angle. Thus this particular beam will be eliminated by total internal reflection. Fig. 18.9 shows a properly cut calcite crystal in which a layer of Canada Balsam has been introduced so that the ordinary ray undergoes total internal reflection. The extraordinary component passes through and the beam emerging from the crystal is linearly polarized.

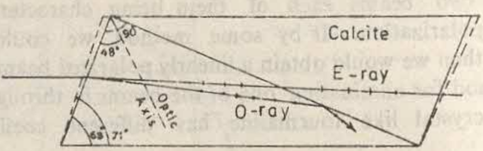


Fig. 18.9 The Nicol prism. The dashed outline corresponds to the natural crystal which is cut in such a way that the ordinary ray undergoes total internal reflection at the Canada balsam layer.

18.2.4 Polarization by Scattering

If an unpolarized beam is allowed to fall on a gas, then the beam scattered at 90° to the incident beam is linearly polarized. This follows from the fact that the waves propagating in the y-direction are produced by the x-component of the dipole oscillations (see Fig. 18.10). The y-component of the dipole oscillations will produce no field in the y-direction (see Sec. 19.4). It may be of interest to mention that it was through scattering experiments that Barkla could establish the transverse character o X-rays. Clearly, if the incident beam is linearly polarized with its electri vector along the x-direction. then there will be no scattered light alon the x-axis. As such, one can carry out an analysis of a scattered wav by allowing it to undergo a further scattering [see Fig. 18.10(b)].

18.3 Malus' Law

Let us consider a polarizer which has a pass-axis parallel to the y-axis*; i.e. if a beam propagating in the z-direction is incident on the polarizer, then the electric vector associated with the emergent wave will oscillate along the y-axis. Consider the incidence of a linearly polarized beam with its electric vector oscillating along a direction which makes an angle θ with the y-axis (see Fig. 18.11). If the amplitude of the electric field is E_0, then the amplitude of the wave emerging from the polaroid will be $E_0 \cos \theta$ and thus the intensity of the emerging beam will be given by

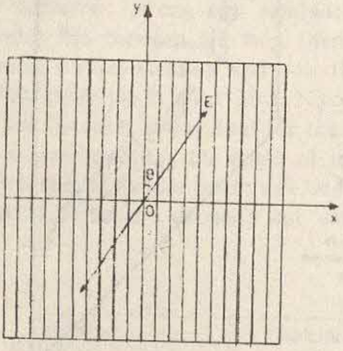


Fig. 18.11 Incidence of a linearly polarized plane wave on a polaroid whose pass-axis is along the y-direction.

$$I = I_0 \cos^2 \theta \tag{18.3-1}$$

where I_0 represents the intensity of the incident beam. Eq. (18.3-1) represents the Malus' Law. Thus, if a linearly polarized beam is incident on a polaroid and if the polaroid is rotated about the z-axis, then the intensity of the emergent wave will vary according to the above law. For example, if the polaroid shown in Fig. 18.11 is rotated in the clockwise direction, then the intensity will increase till the pass-axis is parallel to **OE** a further rotation will result in a decrease in intensity till the pass-axis is perpendicular to **OE**, where the intensity will be almost zero. If we further rotate it, it will pass through a maximum and again a minimum before it reaches its original position.

18.4 Superposition of Two Disturbances

Let us consider the propagation of two linearly polarized electromagnetic waves (both propagating along the z-axis) with their electric vectors oscillating along the x-axis. The electric fields associated with the waves can be written in the form

---

*It should be noted that if the polarizer is a polaroid, then for the pass axis to be along the y-direction, the long chain molecules must be aligned along the x-axis.

$$\mathbf{E}_1 = \hat{\mathbf{x}} a_1 \cos(kz - \omega t + \theta_1) \tag{18.4-1}$$

$$\mathbf{E}_2 = \hat{\mathbf{x}} a_2 \cos(kz - \omega t + \theta_2) \tag{18.4-2}$$

where a_1 and a_2 represent the amplitudes of the waves: $\hat{\mathbf{x}}$ represents the unit vector along the x-axis and θ_1 and θ_2 are phase constants. The resultant of these two waves would be given by

$$\mathbf{E} = \mathbf{E}_1 + \mathbf{E}_2 \tag{18.4-3}$$

which can always be written in the form

$$\mathbf{E} = \hat{\mathbf{x}} a \cos(kz - \omega t + \theta) \tag{18.4-4}$$

where

$$a = [a_1^2 + a_2^2 + 2a_1 a_2 \cos(\theta_1 - \theta_2)]^{1/2} \tag{18.4-5}$$

represents the amplitude of the wave. Eq. (18.4-4) tells us that the resultant is also a linearly polarized wave with its electric vector oscillating along the same axis.

We next consider the superposition of two linearly polarized electromagnetic waves (both propagating along the z-axis) but with their electric vectors oscillating along two mutually perpendicular directions. Thus, we may have

$$\mathbf{E}_1 = \hat{\mathbf{x}} a_1 \cos(kz - \omega t) \tag{18.4-6}$$

$$\mathbf{E}_2 = \hat{\mathbf{y}} a_2 \cos(kz - \omega t + \theta) \tag{18.4-7}$$

For $\theta = n\pi$, the resultant will also be a linearly polarized wave with its electric vector oscillating along a direction making a certain angle with the x-axis; this angle will depend on the relative values of a_1 and a_2.

In order to find the state of polarization of the resultant field, we consider the time variation of the resultant electric field at an arbitrary plane perpendicular to the z-axis which we may, without any loss of generality, assume to be $z = 0$. If E_x and E_y represent the x- and y-components of the resultant field $\mathbf{E} (= \mathbf{E}_1 + \mathbf{E}_2)$, then

$$E_x = a_1 \cos \omega t \tag{18.4-8}$$

and

$$E_y = a_2 \cos(\omega t - \theta) \tag{18.4-9}$$

where we have used Eqs. (18.4-6) and (18.4-7) with $z =$ [illegible]. For $\theta = n\pi$, the above equations simplify to

$$E_x = a_1 \cos \omega t \quad \text{and} \quad E_y = (-1)^n a_2 \cos \omega t \tag{18.4-10}$$

from which one obtains

$$\frac{E_y}{E_x} = \pm\frac{a_2}{a_1} \quad \text{(independent of } t\text{).} \qquad (18.4\text{-}11)$$

where the upper and lower signs correspond to n even and n odd respectively. In the E_x-E_y plane, Eq. (18.4-11) represents a straight line; the angle (φ) that this line makes with the E_x axis depends on the ratio a_2/a_1. In fact

$$\varphi = \tan^{-1}\left(\pm\frac{a_2}{a_1}\right) \qquad (18.4\text{-}12)$$

The condition $\theta = n\pi$ implies that the two vibrations are either in phase ($n = 0, 2, 4, \ldots$) or out of phase ($n = 1, 3, 5, \ldots$). Thus, the superposition of two linearly polarized electromagnetic waves with their electric fields at right angles to each other and oscillating in phase, is again a linearly polarized wave with its electric vector, in general, oscillating in a direction which is different from the fields of either of the waves. Fig. 18.12 is a plot of the resultant field corresponding to Eq. (18.4-10) for various values of a_2/a_1. The tip of the electric vector oscillates (with angular frequency ω)

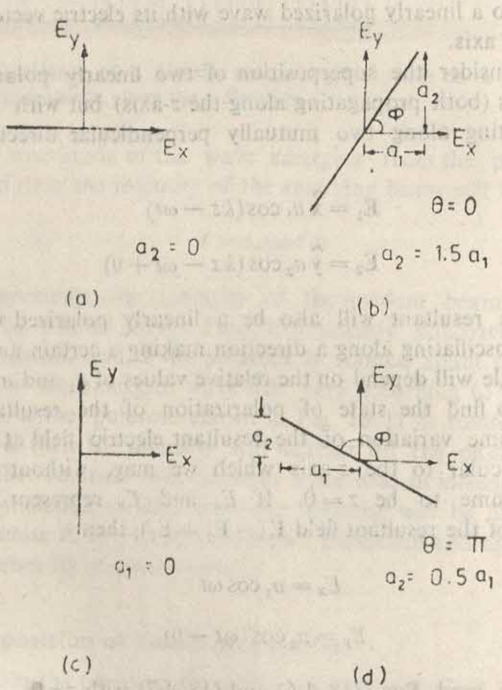


Fig. 18.12 The superposition of two linearly polarized waves with their electric fields oscillating in phase along the x-axis and the y-axis, is again a linearly polarized wave with its electric vector oscillating in a direction making an angle φ with the x-axis.

along the thick lines shown in the figure. The equation of the straight line is given by Eq. (18.4-11).

For $\theta \neq n\pi$, ($n = 0, 1, 2 \ldots$) the resultant electric vector does not, in general, oscillate along a straight line. We first consider the simple case corresponding to $\theta = \pi/2$ with $a_1 = a_2$. Thus,

$$E_x = a_1 \cos \omega t \tag{18.4-13}$$

$$E_y = a_1 \sin \omega t \tag{18.4-14}$$

If we plot the time variation of the resultant electric vector whose x and y-components are given by Eqs. (18.4-13) and (18.4-14), one would find

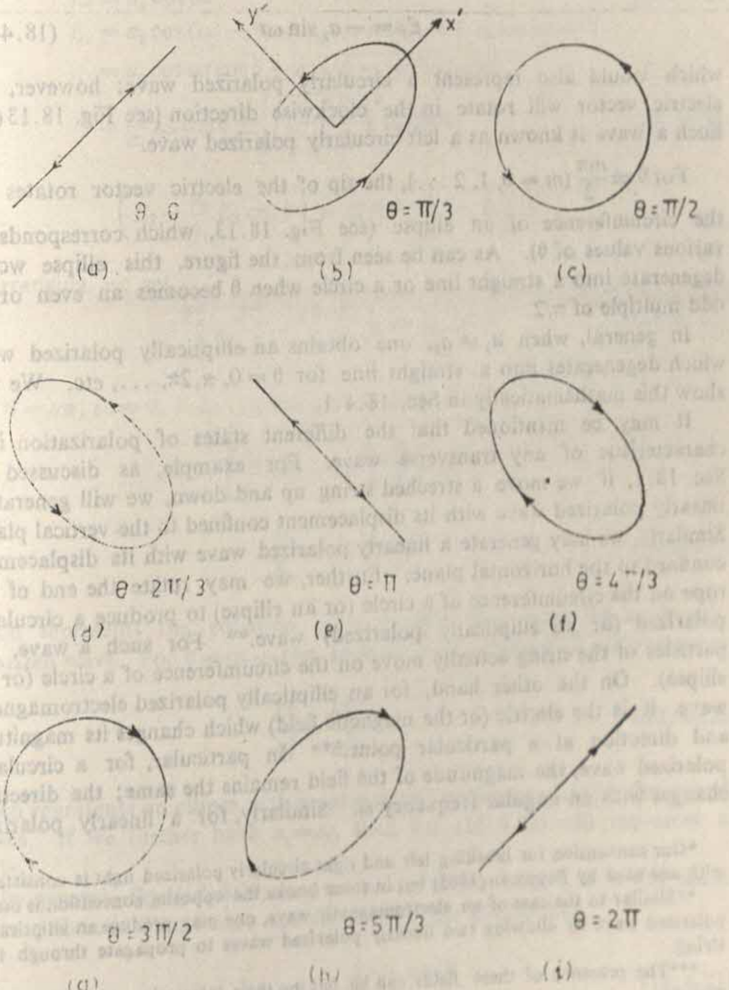


Fig. 18.13 States of polarization for various values of θ corresponding to $a_1 = a_2$ [see Eqs. (18.4-8) and (18.4-9)]. For example, (c) and (g) correspond to right circularly and left circularly polarized light respectively.

that the tip of the electric vector rotates on the circumference of a circle (of radius a_1) in the anti-clockwise direction [see Fig. 18.13 (c)]. Such a wave is known as a right circularly* polarized wave. That the tip of the resultant electric vector should lie on the circumference of a circle is also obvious from the fact that

$$E_x^2 + E_y^2 = a_1^2 \quad \text{(independent of } t\text{)} \qquad (18.4\text{-}15)$$

For $\theta = 3\pi/2$, one would have had

$$E_x = a_1 \cos \omega t \qquad (18.4\text{-}16)$$

$$E_y = -a_1 \sin \omega t \qquad (18.4\text{-}17)$$

which would also represent a circularly polarized wave; however, the electric vector will rotate in the clockwise direction [see Fig. 18.13(g)]. Such a wave is known as a left circularly polarized wave.

For $\theta \neq \frac{m\pi}{2}$ $(m = 0, 1, 2 \ldots)$, the tip of the electric vector rotates on the circumference of an ellipse (see Fig. 18.13, which corresponds to various values of θ). As can be seen from the figure, this ellipse would degenerate into a straight line or a circle when θ becomes an even or an odd multiple of $\pi/2$.

In general, when $a_1 \neq a_2$, one obtains an elliptically polarized wave which degenerates into a straight line for $\theta = 0, \pi, 2\pi, \ldots$, etc. We will show this mathematically in Sec. 18.4.1.

It may be mentioned that the different states of polarization is a characteristic of any transverse wave. For example, as discussed in Sec. 18.1, if we move a streched string up and down, we will generate a linearly polarized wave with its displacement confined to the vertical plane. Similarly, we may generate a linearly polarized wave with its displacement confined to the horizontal plane. Further, we may rotate the end of the rope on the circumference of a circle (or an ellipse) to produce a circularly polarized (or an elliptically polarized) wave.** For such a wave, the particles of the string actually move on the circumference of a circle (or an ellipse). On the other hand, for an elliptically polarized electromagnetic wave, it is the electric (or the magnetic field) which changes its magnitude and direction at a particular point.*** In particular, for a circularly polarized wave, the magnitude of the field remains the same; the direction changes with an angular frequency ω. Similarly, for a linearly polarized wave, the direction of the field does not change; it is the magnitude which keeps on oscillating about the zero value with the angular frequency of the wave.

---

*Our convention for labelling left and right circularly polarized light is consistent with one used by Feynman (1963) but in some books the opposite convention is used.

**Similar to the case of an electromagnetic wave, one may produce an elliptically polarized wave by allowing two linearly polarized waves to propagate through the string.

***The presence of these fields can be felt by their interaction with a charged particle.

18.4.1 The Mathematical Analysis

In this section, we will explicitly show that if we eliminate t from Eqs. (18.4-8) and (18.4-9) we will obtain the equation for an ellipse. We rewrite these equations

$$E_x = a_1 \cos \omega t \tag{18.4-18}$$

$$\begin{aligned} E_y &= a_2 \cos(\omega t - \theta) = a_2 \cos \omega t \cos \theta + a_2 \sin \omega t \sin \theta \\ &= a_2 \cos \omega t \cos \theta \pm a_2 \sqrt{1 - \cos^2 \omega t}\, \sin \theta \\ &= \frac{a_2}{a_1} E_x \cos \theta \pm \sqrt{1 - \left(\frac{E_x}{a_1}\right)^2}\, a_2 \sin \theta \end{aligned} \tag{18.4-19}$$

or

$$\left(E_y - \frac{a_2}{a_1} E_x \cos \theta\right)^2 = a_2^2 \sin^2 \theta \left[1 - \left(\frac{E_x}{a_1}\right)^2\right]$$

Rearranging, we get

$$\frac{E_x^2}{a_1^2} + \frac{E_y^2}{a_2^2} - \frac{2E_x E_y}{a_1 a_2} \cos \theta = \sin^2 \theta \tag{18.4-20}$$

For $\theta = n\pi$; $(n = 0, 1, 2 \ldots)$, Eq. (18.4-20) becomes

$$\left[\frac{E_x}{a_1} - (-1)^n \frac{E_y}{a_2}\right]^2 = 0$$

i.e.

$$\frac{E_y}{E_x} = (-1)^n \frac{a_2}{a_1} \tag{18.4-21}$$

which represents the equation of a straight line, implying a linearly polarized wave. For $\theta = (n + \frac{1}{2})\pi$, $(n = 0, 1, 2, \ldots)$, Eq. (18.4-20) becomes

$$\frac{E_x^2}{a_1^2} + \frac{E_y^2}{a_2^2} = 1 \tag{18.4-22}$$

which represents an ellipse with its minor and major axes along the x-and y-axes. If we further have $a_1 = a_2$, then Eq. (18.4-22) will represent a circle.

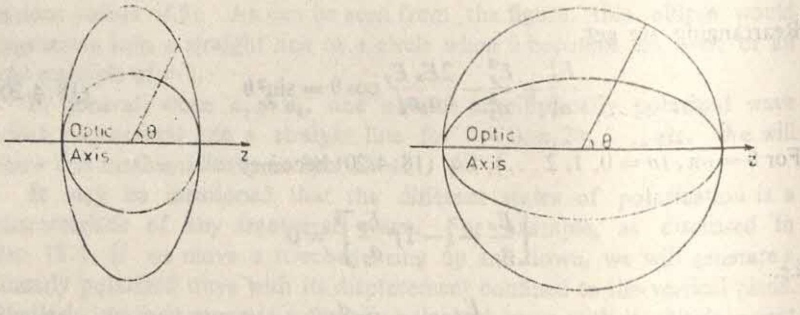
In general, Eq. (18.4-20) represents an ellipse and if φ is the angle between the coordinate axes and the axes of the ellipse then

$$\tan 2\varphi = \frac{2a_1 a_2 \cos \theta}{a_1^2 - a_2^2} \tag{18.4-23}$$

(see Problem 16).

18.5 The Phenomenon of Double Refraction

If we place a calcite crystal on a point marked in a piece of paper, we will, in general, see two images of the point. The appearance of the two dots is due to the phenomenon of double refraction. In general, a ray entering such a crystal will split up into two rays and for a crystal like calcite, one of the rays will follow the Snell's law of refraction and the other will not. The former is referred to as the ordinary ray and the latter as the extraordinary ray. The velocity of the ordinary ray is the same in all directions whereas the velocity of the extraordinary ray is different in different directions.* Along a particular direction, the two velocities are equal; this direction is known as the optic axis of the crystal.** Indeed, if we imagine a point source imbedded in a crystal like calcite, then we will have two different sets of wavefronts, one corresponding to the ordinary ray which is spherical and the other corresponding to the extraordinary ray which is an ellipsoid of revolution. An ellipsoid of revolution is a surface obtained by rotating an ellipse about either the major or the minor axis; this axis of rotation is along the optic axis (see Fig. 18.14). If the ellipsoid of revolution lies completely



Negative Crystal

(a)

Positive Crystal

(b)

Fig. 18.14 The ellipsoid of revolution (which corresponds to the extraordinary ray) lies outside the sphere (which corresponds to the ordinary ray) in a negative crystal as shown in (a); in a positive crystal, the converse is true-see (b).

outside the sphere (i.e. if the velocity of the extraordinary ray is everywhere greater than the ordinary ray [see Fig. 18.14(a)] then the crystal is known as a negative crystal. This is indeed the case for calcite.

---

*A substance which exhibits different properties in different directions is called an anisotropic substance.

**In a crystal like calcite, the two waves have the same speed only along one direction (which is the optic axis); such crystals are known as uniaxial crystals. In general, there may be two directions along which the two waves have the same speed; such crystals are known as biaxial crystals—see Sec. 18.8.

On the other hand, if the ellipsoid of revolution lies completely inside the sphere, i.e. if the velocity of the extraordinary ray is everywhere less than the ordinary ray [see Fig. 18.14(b)], then the crystal is known as a positive crystal which is the case for quartz. In either case, the velocities of the ordinary and the extraordinary rays are given by the following equations:*

$$v_{ro} = v_x \qquad \text{(ordinary ray)} \qquad (18.5\text{-}1)$$

$$\frac{1}{v_{re}^2} = \frac{\sin^2\theta}{v_z^2} + \frac{\cos^2\theta}{v_x^2} \qquad \text{(extraordinary ray)} \qquad (18.5\text{-}2)$$

where θ is the angle that the ray makes with the optic axis, which we will assume to be parallel to the z-axis; v_x and v_z are constants of the crystal.** If we plot v_{re} as a function of θ we will obtain an ellipsoid of revolution; on the other hand, since v_{ro} is independent of θ, if we plot v_{ro} (as a function of θ), we will obtain a sphere. Along the optic axis, $\theta = 0$ and $v_{ro} = v_{re}$. When $v_z < v_x$, v_r, will be greater than v_{ro} (i.e. it will be negative crystal) and conversely.

Both the ordinary and the extraordinary rays are linearly polarized, the vibration corresponding to the ordinary ray is always at right angles to the direction of the ray as well as to the optic axis. On the other hand, for the extraordinary ray, the direction of vibration is at right angles to the propagation vector **k** and lies in the plane containing the optic axis and the ray direction.*** In this section, we will consider the refraction of a plane electromagnetic wave incident on a negative crystal like calcite. Similar analysis can be carried out for positive crystals also.

A. *Normal incidence*

We first consider the refraction of a plane electromagnetic wave by a uniaxial crystal. The direction of the optic axis is shown as a dashed line in Fig. 18.15. In order to determine the ordinary ray we draw a sphere of radius v_x from the point B. Similarly, we draw another sphere from the point D. The common tangent plane to these spheres is shown as OO′ which represents the wavefront corresponding to the ordinary refracted ray. In order to determine the extraordinary ray, we draw an ellipse (centred at the point B) with its minor axis along the optic axis. The ellipsoid of revolution is obtained by rotating the ellipse about the optic axis. Similarly, we draw another ellipsoid of revolution from the point D. The common tangent plane to these ellipsoids (which will be perpendicular to **k**) is shown as EE' in Fig. 18.15. If we join the point B to the point of contact O, then

---

*These relations have been derived in Sec. 18.8.

**It may be pointed out that v_x and v_z are not the velocities of the extraordinary ray when it propagates along the x-and z-axis, in fact it is just the reverse; v_x and v_z are the velocities of the extraordinary ray when it propagates along the z-and x-axes respectively.

***For a plane wave, the space and time dependence of the vectors **E**, **D**, **B** and **H** are of the form exp $[i(\mathbf{k}\cdot\mathbf{r}-\omega t)]$, where **k** denotes the propagation vector and represents the direction normal to the phase fronts. We will show in Sec. 18.8 that **D** is always at right angles to **k** and hence the direction of **D** is chosen as the direction of "vibrations".

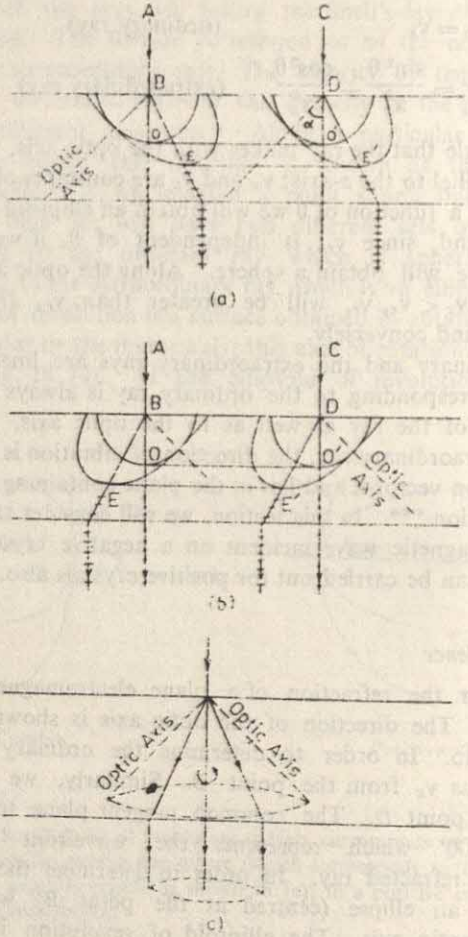


Fig. 18.15 (a) and (b) The refraction of a plane wave incident on a negative crystal whose optic axis is along the dashed line. (c) For a ray incident normally, if the crystal is rotated about a line which is normal to the surface, the extraordinary ray will rotate on the periphery of a cone.

corresponding to the incident ray AB, the direction of the ordinary ray will be along BO. Similarly, if we join the point B to the point of contact E (between the ellipsoid of revolution and the tangent plane EE'), then corresponding to the incident ray AB, the direction of the extraordinary ray* will be along BE.

Obviously, if we have a different direction of the optic axis [see Fig. 18.15(b)] then although the direction of the ordinary ray will remain the same, the extraordinary ray will propagate in a different direction. Thus if a ray is incident normally on a calcite crystal and if the crystal is rotated about the normal, then the extraordinary ray will also rotate (about the normal) on the periphery of a cone; each time the ray will lie in the plane containing the normal and the optic axis [see Fig. 18.15(c)].

As mentioned earlier, the direction of vibrations for the ordinary ray is normal to the optic axis and the ordinary ray; as such, in this case, they will be normal to the plane of the paper. Therefore, the direction of these vibrations have been shown as dots in Fig. 18.15. Similarly, since the direction of vibrations for the extraordinary ray is perpendicular to **k** and lies in the plane containing the extraordinary ray and the optic axis, they are along the small straight lines drawn on the extraordinary ray in Fig. 18.15.

Thus, an incident ray will split up into two rays propagating in different directions and when they leave the crystal, we will obtain two linearly polarized beams.

In the above case, we have assumed the optic axis to make an arbitrary angle α with the normal to the surface. In the special cases of $\alpha = 0$ and $\alpha = \pi/2$, the ordinary and the extraordinary rays travel along the same direction as shown in Figs. 18.16(a) and 18.16(b). Fig. 18.16(c) corresponds to the case when the optic axis is normal to the plane of the paper; and as such, the section of the extraordinary wavefront in the plane of the paper will be a circle. Once again, both the ordinary and the extraordinary rays travel along the same direction. It may be mentioned that Figs. 18.16(a) and 18.16(c) indeed correspond to the same configuration; in fact, they represent two different cross-sections of the same set of spherical and ellipsoidal wavefronts.

Notice that in the configuration shown in Fig. 18.16(a), although both the rays travel in the same direction, they propagate with different velocities. This phenomenon is used in the fabrication of quarter and half-wave plates (see Sec. 18.6). On the other hand, in the configuration shown in Fig. 18.16(b), both the rays not only travel in the same direction but they also propagate with the same velocity.

---

*The ray refractive index corresponding to the extraordinary ray (n_{re}) will be given by

$$n_{re} = \frac{c}{v_{re}} = [n_{\parallel}^2 \cos^2\theta + n_{\perp}^2 \sin^2\theta]^{1/2}$$

where $n_{\parallel} = c/v_x$, $n_{\perp} = c/v_z$; v_x and v_z are the velocities of the extraordinary ray when it propagates parallel and perpendicular to the optic axis. If one starts with the above equation and uses Fermat's principle to obtain the refracted ray, the results will be consistent with the ones obtained in this section (see Sec. 2.4).

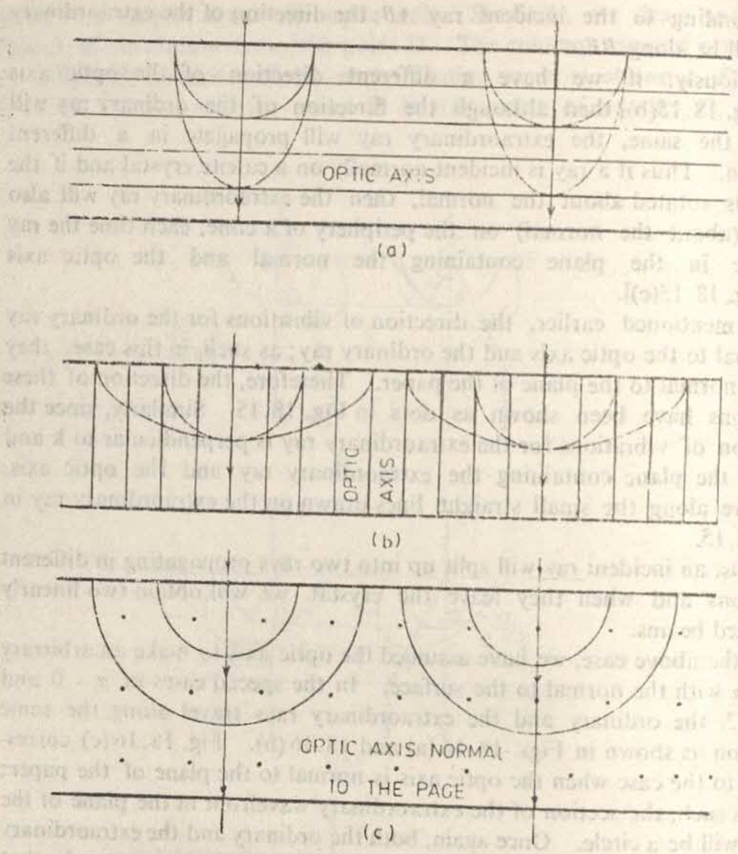


Fig. 18.16 Propagation of a plane wave incident normally on a negative uniaxial crystal. In (a) and (b) the optic axis is shown as parallel straight lines and in (c) the optic axis is perpendicular to the plane of the figure and is shown as dots. In each case, the extraordinary and the ordinary rays travel in the same direction.

B. *Oblique incidence*

We next consider the case of a plane wave incident obliquely on a negative uniaxial crystal [see Fig. 18.17 (a)]. Once again we use Huygens' principle to determine the shape of the refracted wavefronts. Let *BD* represent the incident wavefront. If the time taken for the disturbance to reach the point *F* from the point *D* is t, then with *B* as centre we draw a sphere of radius $v_x t$ and an ellipsoid of revolution of semi-minor and semi-major axes $v_x t$ and $v_z t$ respectively; the semi-minor axis is along the optic axis. From the point *F* we draw tangent planes *FO* and *FE* to the sphere and the ellipsoid of revolution respectively. These planes would represent the refracted wavefronts corresponding to the ordinary and the extraordinary rays respectively. If the points of contacts are O and E, then the ordinary and extraordinary refracted rays will propagate along BO and BE respectively.* The direction of vibrations of these rays are shown by dots and

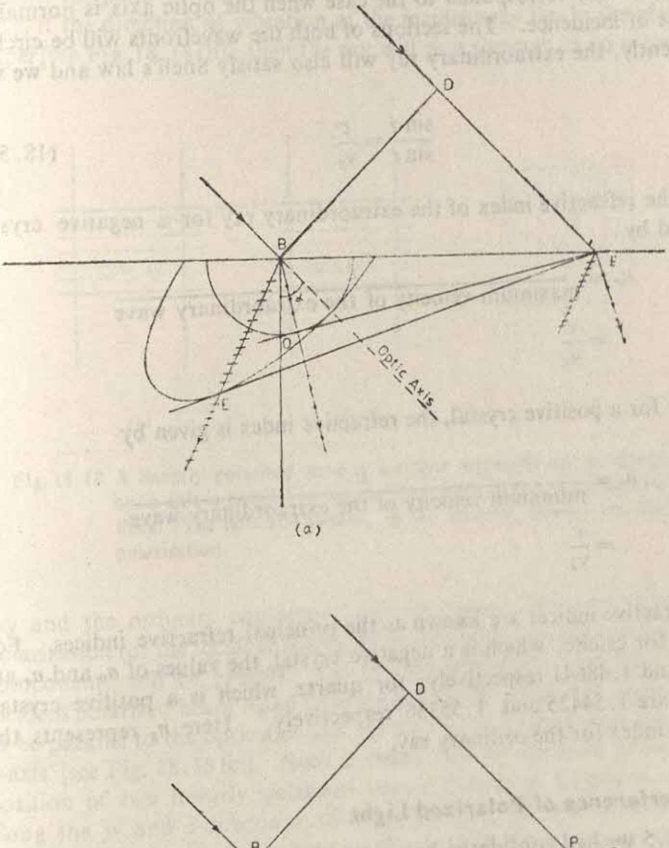


Fig. 18.17 Refraction of a plane wave incident obliquely on a negative uniaxial crystal. In (a) the direction of the optic axis is along the dashed line. In (b), the optic axis is perpendicular to the plane of the paper.

---

*This can also be shown by using Fermat's principle—See Sec. 2.4.

small lines respectively and are obtained by using the general rules discussed earlier. The shape of the refracted wavefronts corresponding to the particular case of $\alpha = 0$ and $\alpha = \pi/2$ can be obtained very easily.

Fig. 18.17(b) corresponds to the case when the optic axis is normal to the plane of incidence. The sections of both the wavefronts will be circles; consequently, the extraordinary ray will also satisfy Snell's law and we will have

$$\frac{\sin i}{\sin r} = \frac{c}{v_z} \tag{18.5-3}$$

Indeed the refractive index of the extraordinary ray for a negative crystal is defined by

$$n_e = \frac{c}{\text{maximum velocity of the extraordinary wave}}$$

$$= \frac{c}{v_z}$$

Similarly for a positive crystal, the refractive index is given by

$$n_e = \frac{c}{\text{minimum velocity of the extraordinary wave}}$$

$$= \frac{c}{v_z}$$

These refractive indices are known as the principal refractive indices. For example, for calcite, which is a negative crystal, the values of n_0 and n_e are 1.65836 and 1.48641 respectively; for quartz, which is a positive crystal, n_0 and n_e are 1.54425 and 1.55336 respectively.* Here n_0 represents the refractive index for the ordinary ray.

18.6 Interference of Polarized Light

In Sec. 18.5 we had considered how a ray (incident on a doubly refracting crystal) splits up into two rays each characterized by a certain state of polarization. The direction of vibration associated with the ordinary ray is at right angles to the direction of the ray as well as to the optic axis; and the direction of vibration for the extraordinary ray lies in the plane containing the optic axis and the ray direction and is perpendicular to **k**. The considerations in Sec. 18.5 assumed that the incident beam was unpolarized. In this section, we will consider the normal incidence of a plane polarized beam and study the state of polarization of the beam emerging from the crystal.

We assume the optic axis to be parallel to the surface of the crystal

---

*These values correspond to 18°C for $\lambda = 5892.90$ Å; quoted from Jenkins and White (1976).

[see Fig. 18.16(a)]. If the direction of vibration of the incident ray is perpendicular to the plane of the paper [see Fig. 18.18(a)], then the ray will pass through as an ordinary ray and the extraordinary component will be absent; the plane of the paper is assumed to contain the optic axis. Similarly, if the direction of vibration of the incident ray lies in the plane of the paper [see Fig. 18.18(b)], then the ray will pass through as an extraordinary ray and the ordinary component will be absent. For any other state of polarization (of the incident beam), both the extraordinary and the ordinary components will be present. Let the electric vector associated with the incident polarized beam make an angle φ with the z-axis; the z-axis is chosen to be parallel to the optic axis and the direction of propagation is along the x-axis [see Fig. 18.18(c)]. Such a beam can be assumed to be a superposition of two linearly polarized beams (vibrating in phase), polarized along the y- and z-directions with amplitudes $E \sin \varphi$ and $E \cos \varphi$ respectively. The z-component (whose amplitude is $E \cos \varphi$) passes through as an extraordinary beam propagating with speed c/n_e (see Sec. 18.5). The y-component (whose amplitude is $E \sin \varphi$) passes through as an ordinary beam propagating with speed c/n_0. Since $n_e \neq n_0$, the two beams will propagate with different velocities and, as such, when they come out of the crystal, they will not be in phase. Consequently, the emergent beam (which will be a superposition of these two beams) will be, in general, elliptically polarized.

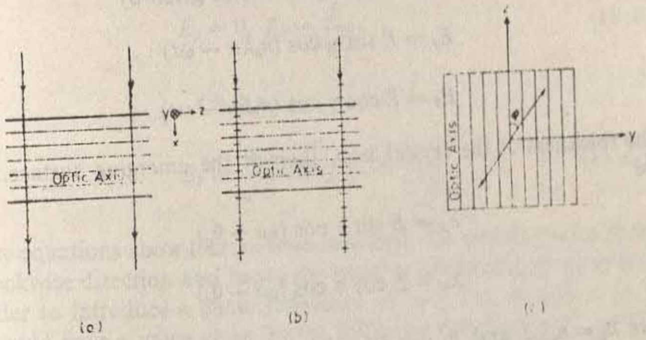


Fig. 18.18 A linearly polarized wave is incident normally on a crystal whose optic axis is parallel to the surface which is shown as parallel straight lines. The dots and dashes, in (a) and (b), denote the direction of polarization.

Let the plane $x = 0$ represent the surface of the crystal on which the beam is incident. The y- and z-components of the incident beam can be written in the form

$$E_y = E \sin \varphi \cos (kx - \omega t) \tag{18.6-1}$$

$$E_z = E \cos \varphi \cos (kx - \omega t)$$

where $k\,(=\omega/c)$ represents the free-space wave number. Thus, at $x=0$, we will have

$$\left.\begin{aligned} E_y(x=0) &= E\sin\varphi\cos\omega t \\ E_z(x=0) &= E\cos\varphi\cos\omega t \end{aligned}\right\} \qquad (18.6\text{-}2)$$

Inside the crystal, the two components will be given by

$$\begin{aligned} E_y &= E\sin\varphi\cos(n_0kx-\omega t) \\ E_z &= E\cos\varphi\cos(n_ekx-\omega t) \end{aligned} \qquad (18.6\text{-}3)$$

If the thickness of the crystal is d, then at the emerging surface, we will have

$$\begin{aligned} E_y &= E\sin\varphi\cos(\omega t-\theta_0) \\ E_z &= E\cos\varphi\cos(\omega t-\theta_e) \end{aligned} \qquad (18.6\text{-}4)$$

where $\theta_0=n_0kd$ and $\theta_e=n_ekd$. By appropriately choosing the instant $t=0$, the components may be rewritten as

$$\begin{aligned} E_y &= E\sin\varphi\cos(\omega t-\theta) \\ E_z &= E\cos\varphi\cos\omega t \end{aligned} \qquad (18.6\text{-}5)$$

where

$$\theta=\theta_0-\theta_e=kd(n_0-n_e)=\frac{\omega}{c}(n_0-n_e)\,d \qquad (18.6\text{-}6)$$

represents the phase difference between the ordinary and the extraordinary beams. Clearly, if the thickness of the crystal is such that $\theta=2\pi, 4\pi, \ldots$ the emergent beam wlll have the same state of polarization as the incident beam [see Fig. 18.13 (a)]; if $\theta=\pi, 3\pi, \ldots$ the emergent beam will again be linearly polarized but the electric vector will be along a direction as shown in Fig 18.13 (e). For any other value of θ the emergent beam will be elliptically polarized; in Fig. 18.13 we have plotted the different states of polarization corresponding to $\varphi=\pi/4$ with $\theta=0, \pi/3, \pi/2, \ldots$ etc.

As an example, let us consider the case when $\varphi=\pi/4$ and $\theta=\pi/2$; i.e. the y- and z-components of the incident wave have equal amplitudes and the crystal introduces a phase difference of $\pi/2$. Thus, for the emergent beam we have

$$\begin{aligned} E_y &= \frac{E}{\sqrt{2}}\sin\omega t \\ E_z &= \frac{E}{\sqrt{2}}\cos\omega t \end{aligned} \qquad (18.6\text{-}7)$$

which represents a circularly polarized wave because

$$E_y^2 + E_z^2 = \frac{E^2}{2} \tag{18.6-8}$$

In order to determine the direction of rotation of the electric vector, we note that at $t = 0$,

$$E_y = 0, \; E_z = \frac{E}{\sqrt{2}} \tag{18.6-9}$$

and at $t = \Delta t$

$$E_y \simeq \frac{E}{\sqrt{2}} \omega \Delta t, \; E_z \simeq \frac{E}{\sqrt{2}} \tag{18.6-10}$$

The above equations show that as time increases, the electric vector rotates in the clockwise direction and hence the beam is left circularly polarized.

In order to introduce a phase difference of $\pi/2$, the thickness of the crystal should have a value given by the following equation:

$$d = \frac{c}{\omega(n_0 - n_e)} \frac{\pi}{2} = \frac{1}{4} \frac{\lambda}{(n_0 - n_e)} \tag{18.6-11}$$

where λ is the free-space wavelength. For calcite

$$n_0 = 1.65836, \quad n_e = 1.48641 \tag{18.6-12}$$

which correspond to $\lambda = 5893$ Å and at 18°C. Substituting these values, we obtain

$$d = \frac{5893 \times 10^{-8}}{4 \times 0.17195} \text{ cm} \simeq 0.000857 \text{ mm} \tag{18.6-13}$$

Thus a calcite crystal of thickness 0.000857 mm with its optic axis parallel to the surface will introduce a phase difference of $\pi/2$ (and hence a path difference of $\lambda/4$) between the ordinary and the extraordinary components at $\lambda = 5893$ Å. Such a plate is known as a quarter wave plate. It should be pointed out that if the thickness is an odd multiple of the above quantity, i.e. if

$$d = (2m + 1) \frac{\lambda}{4(n_0 - n_e)}; \; m = 0, 1, 2, .. \tag{18.6-14}$$

it will still produce a path difference of $\lambda/4$. In the example considered above (i.e. when $\varphi = \pi/4$), it can easily be shown that the emergent beam will be left circularly polarized for $m = 0, 2, 4...$and right circularly polarized for $m = 1, 3, 4...$

For a positive crystal (like quartz), $n_e > n_0$ and Eq. (18.6-5) should be written in the form

$$E_y = E \sin \varphi \cos (\omega t + \theta') \qquad (18.6\text{-}15)$$

$$E_z = E \cos \varphi \cos \omega t$$

where $\theta' = \frac{\omega}{c} d (n_e - n_0)$. For a quarter wave plate

$$d = (2m+1) \frac{\lambda}{4(n_e - n_0)};\ m = 0, 1, 2, \ldots \qquad (18.6\text{-}16)$$

However, in this case, for $\varphi = \pi/4$, the emergent beam will be right circularly polarized for $m = 0, 2, 4\ldots$ and left circularly polarized for $m = 1, 3, 5\ldots$

When the thickness of the crystal is such that it introduces a path difference of $\lambda/2$, then it is referred to as a half-wave plate. Thus, for a half-wave plate, the thickness (for a negative crystal) would be given by

$$d = (2m+1) \frac{\lambda}{2(n_0 - n_e)} \qquad (18.6\text{-}17)$$

Example 18.1 A left circularly polarized beam ($\lambda = 5893$ Å) is incident normally on a calcite crystal (with its optic axis cut parallel to the surface) of thickness 0.005141 mm. What will be the state of polarization of the emergent beam?

Solution. The electric field for the incident beam at $x = 0$ would be

$$E_y = \frac{E}{\sqrt{2}} \sin \omega t$$

$$E_z = \frac{E}{\sqrt{2}} \cos \omega t$$

Now,

$$\theta = \frac{(n_0 - n_e) d \times 2\pi}{\lambda} = \frac{0.17195 \times 0.005141 \times 2\pi}{5893 \times 10^{-7}}$$

$$\simeq 3\pi$$

Thus the emergent wave will be [cf. Eq. (18.6-5)]

$$E_y = \frac{E}{\sqrt{2}} \sin (\omega t - 3\pi) = -\frac{E}{\sqrt{2}} \sin \omega t$$

$$E_z = \frac{E}{\sqrt{2}} \cos \omega t$$

which represents a right circularly polarized beam.

Example 18.2 A left circularly polarized beam ($\lambda = 5893$ Å) is incident on a quartz crystal (with its optic axis cut parallel to the surface) of thickness 0.025 mm. Determine the state of polarization of the emergent beam. Assume n_0 and n_e to be 1.54425 and 1.55336 respectively.

Solution. As in the previous example, the electric field for the incident beam at $x = 0$ would be given by Eq. (18.6-15). Further,

$$\theta' = (n_e - n_0)\frac{2\pi}{\lambda}d = 2\pi\,\frac{0.00911\times 0.025}{5893\times 10^{-7}} \simeq 0.77\,\pi$$

Thus the emergent beam will be

$$E_y = \frac{E}{\sqrt{2}}\sin(\omega t + 0.77\,\pi)$$

$$E_z = \frac{E}{\sqrt{2}}\cos\omega t$$

which will represent a right elliptically polarized light.

18.7 Analysis of Polarized Light

In the previous sections we have seen that a plane wave can be characterized by different states of polarizations, which may be any one of the following: (a) linearly polarized, (b) circularly polarized, (c) elliptically polarized, (d) unpolarized, mixture of linearly polarized and unpolarized, (f) mixture of circularly polarized and unpolarized or (g) mixture of elliptically polarized and unpolarized light. To the naked eye all the states of polarizations will appear to be the same. In this section, we will discuss the procedure for determining the state of polarization of a light beam.

If we introduce a polaroid in the path of the beam and rotate it about the direction of propagation, then either of the following three possibilities can occur:

(i) If there is complete extinction at two positions of the polarizer, then the beam is linearly polarized.

(Ii) If there is no variation of intensity, then the beam is either unpolarized or circularly polarized or a mixture of unpolarized and circularly polarized light. We now put a quarter wave plate on the path of the beam followed by the rotating polaroid. If there is no variation of intensity then the incident beam is unpolarized. If there is complete extinction at two positions, then the beam is circularly polarized (this is due to the fact that a quarter wave plate will transform a circularly polarized light into a linearly polarized light). If there is a variation of intensity (without complete extinction) then the beam is a mixture of unpolarized and circularly polarized light.

(iii) If there is a variation of intensity (without complete extinction) then the beam is either elliptically polarized or a mixture of linearly polarized and unpolarized or a mixture of elliptically polarized and unpolarized light. We now put a quarter wave plate in front of the polaroid with its optic axis parallel to the pass-axis of the polaroid at the position of maximum intensity. The elliptically polarized light will transform to a linearly polarized light and thus if one obtains two positions of the polaroid where complete extinction occurs, then the original beam is elliptically polarized. If complete extinction does not occur, and the position of maximum intensity occurs at the same orientation as before, then the beam is a mixture of unpolarized and linearly polarized light. Finally, if the position of maximum intensity occurs at a different orientation of the polaroid, the beam is a mixture of elliptically polarized and unpolarized light.

18.8 Plane Wave Propagation in Anisotropic Media

In this section, we will discuss the plane wave solutions of Maxwell's equations in an anisotropic medium and will prove the various assumptions made in Sec. 18.5. The difference between an isotropic and an anisotropic medium is in the relationship between the displacement vector **D**, and the electric vector **E**. In an isotropic medium, **D** is in the same direction as **E** and one can write

$$\mathbf{D} = \epsilon \mathbf{E} \tag{18.8-1}$$

where ϵ is the dielectric permittivity of the medium.* On the other hand, in an anisotropic medium **D** is not, in general, in the direction of **E** and the relation between **D** and **E** can be written in the form

$$\left.\begin{aligned} D_x &= \epsilon_{xx} E_x + \epsilon_{xy} E_y + \epsilon_{xz} E_z \\ D_y &= \epsilon_{yx} E_x + \epsilon_{yy} E_y + \epsilon_{yz} E_z \\ D_z &= \epsilon_{zx} E_x + \epsilon_{zy} E_y + \epsilon_{zz} E_z \end{aligned}\right\} \tag{18.8-2}$$

where $\epsilon_{xx}, \epsilon_{xy}, \ldots$ are constants.** One can show that $\epsilon_{xy} = \epsilon_{yx}$, $\epsilon_{xz} = \epsilon_{zx}$ and $\epsilon_{yz} = \epsilon_{zy}$ (see Problem 15). Further, one can always choose a coordinate system in which the above equations reduce to

$$\left.\begin{aligned} D_x &= \epsilon_x E_x \\ D_y &= \epsilon_y E_y \\ D_z &= \epsilon_z E_z \end{aligned}\right\} \tag{18.8-3}$$

where ϵ_x, ϵ_y and ϵ_z are known as the principal dielectric constants.

Let us consider the propagation of a plane electromagnetic wave; for such a wave the vectors **E**, **H**, **B** and **D** would be proportional to

$$e^{i(\mathbf{k}.\mathbf{r}-\omega t)} \tag{18.8-4}$$

---

*The displacement vector **D** is defined in Sec. 19.9 and the plane wave solutions in an isotropic medium are discussed in Sec. 19.2.

**Eq. (18.8-2) is usually written in the following compact form

$$\mathbf{D} = \boldsymbol{\epsilon}\,\mathbf{E}$$

where $\boldsymbol{\epsilon}$ is a second rank tensor.

where **k** represents the propagation vector and ω, the frequency of the wave (the proportionality constant will be independent of space and time coordinates). The phase velocity v_p, and the refractive index n are defined by the following relations:

$$v_p = \frac{\omega}{k} = \frac{c}{n} \tag{18.8-5}$$

In the absence of any currents (i.e. $\mathbf{J} = 0$) Maxwell's curl equations [see Eqs. (19.1-7) and (19.1-8)] become

$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t} = i\omega\, \mathbf{B} = i\omega\mu_0 \mathbf{H} \tag{18.8-6}$$

and

$$\nabla \times \mathbf{H} = \frac{\partial \mathbf{D}}{\partial t} = -\, i\omega\, \mathbf{D} \tag{18.8-7}$$

where we have assumed the medium to be non-magnetic (i.e. $\mathbf{B} = \mu_0 \mathbf{H}$). Now, if

$$\mathbf{E} = \mathbf{E}_0 \exp\left[i(\mathbf{k}.\mathbf{r} - \omega t)\right] \tag{18.8-8}$$

then

$$(\nabla \times \mathbf{E})_x = \left(\frac{\partial \mathbf{E}_z}{\partial y} - \frac{\partial \mathbf{E}_y}{\partial z}\right) = (i\, k_y\, E_{0z} - i\, k_z\, E_{0y})\, e^{i(\mathbf{k}.\mathbf{r} - \omega t)}$$

$$= i\,(\mathbf{k} \times \mathbf{E})_x$$

Thus

$$\nabla \times \mathbf{E} = i(\mathbf{k} \times \mathbf{E}) = i\omega\mu_0 \mathbf{H} \Rightarrow \mathbf{H} = \frac{\mathbf{k} \times \mathbf{E}}{\omega\mu_0} \tag{18.8-9}$$

and

$$\nabla \times \mathbf{H} = i(\mathbf{k} \times \mathbf{H}) = -\, i\omega\, \mathbf{D} \Rightarrow \mathbf{D} = \frac{\mathbf{H} \times \mathbf{k}}{\omega} \tag{18.8-10}$$

Eqs. (18.8-9) and (18.8-10) show that **H** is at right angles to **k**, **E** and **D**; thus **k**, **E** and **D** must lie in the same plane. Further **D** is at right angles to **k** (see Fig. 18.19). Substituting for **H** in Eq. (18.8-10) we get

$$\mathbf{D} = \frac{1}{\mu_0 \omega^2}\left[(\mathbf{k} \times \mathbf{E}) \times \mathbf{k}\right]$$

$$= \frac{1}{\mu_0 \omega^2}\left[(\mathbf{k}.\mathbf{k})\, \mathbf{E} - (\mathbf{k}.\mathbf{E})\, \mathbf{k}\right] \tag{18.8-11}$$

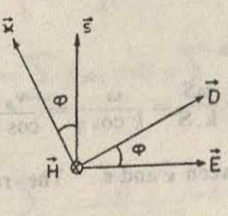


Fig. 18.19 In an anisotropic medium, $\mathbf{D}.\kappa = 0$, $\mathbf{E}.\mathbf{s} = 0$ and **H** is normal to **E**, **D**, **s** and κ.

where we have used the vector identity

$$(\mathbf{A}\times\mathbf{B})\times\mathbf{C}=(\mathbf{A}.\mathbf{C})\,\mathbf{B}-(\mathbf{B}.\mathbf{C})\,\mathbf{A} \qquad (18.8\text{-}12)$$

Thus

$$\mathbf{D}=\frac{k^2}{\mu_0\omega^2}\,[\mathbf{E}-(\boldsymbol{\kappa}.\mathbf{E})\,\boldsymbol{\kappa}]=\frac{k^2}{\mu_0\omega^2}\,\mathbf{E}_\perp \qquad (18.8\text{-}13)$$

where

$$\boldsymbol{\kappa}=\frac{\mathbf{k}}{k} \qquad (18.8\text{-}14)$$

represents the unit vector along **k**, and $\mathbf{E}_\perp$ represents the component of **E** perpendicular to **k** (and hence along **D**, see Fig. 18.19). The direction of energy propagation (or the ray propagation) is along the Poynting vector **S** which is given by

$$\mathbf{S}=\mathbf{E}\times\mathbf{H} \qquad (18.8\text{-}15)$$

Clearly, **S** is at right angles to **E** and **H** and will, therefore, be in the same plane containing the vectors **k**, **E** and **D** (see Fig. 18.19). As can be seen from the figure, the direction of propagation of the wave ($\boldsymbol{\kappa}$) is not along the direction of energy propagation (**s**), where **s** is the unit vector along **S**. The ray velocity (or the energy transmission velocity) v_r is defined as

$$v_r=\frac{S}{u} \qquad (18.8\text{-}16)$$

where u is the energy density. Now,

$$u=\tfrac{1}{2}\,(\mathbf{D}.\mathbf{E}+\mathbf{B}.\mathbf{H})=\tfrac{1}{2}\,(\mathbf{D}.\mathbf{E}+\mu_0\mathbf{H}.\mathbf{H}) \qquad (18.8\text{-}17)$$

[see Eq. (19.4-6)]. Substituting for **H** and **D** from Eqs. (18.8-9) and (18.8-10), we obtain

$$\begin{aligned}u&=\frac{1}{2\omega}\,[(\mathbf{H}\times\mathbf{k}).\mathbf{E}+(\mathbf{k}\times\mathbf{E}).\mathbf{H}]\\&=\frac{1}{2\omega}\,[\mathbf{k}.(\mathbf{E}\times\mathbf{H})+\mathbf{k}.(\mathbf{E}\times\mathbf{H})]\\&=\frac{1}{\omega}\,\mathbf{k}.\mathbf{S}\end{aligned} \qquad (18.8\text{-}18)$$

Thus Eq. (18.8-16) becomes

$$v_r=\frac{\omega S}{\mathbf{k}.\mathbf{S}}=\frac{\omega}{k\cos\varphi}=\frac{v_p}{\cos\varphi} \qquad (18.8\text{-}19)$$

where φ is the angle between $\boldsymbol{\kappa}$ and **s**. The ray refractive index n_r is defined as

$$n_r=\frac{c}{v_r}=\frac{c}{v_p}\cos\varphi=n\cos\varphi \qquad (18.8\text{-}20)$$

In order to express **E** in terms of **D**, we note that $(\mathbf{D}.\hat{\mathbf{E}})\,\hat{\mathbf{E}}$ represents the component of **D** along **E** (see Fig. 18.19); here $\hat{\mathbf{E}}$ represents the unit vector along **E**. Now, the component of **D** along **E** is also $[\mathbf{D}-\mathbf{s}\,(\mathbf{D}.\mathbf{s})]$. Thus

$$\mathbf{D}-\mathbf{s}\,(\mathbf{D}.\mathbf{s}) = (\mathbf{D}.\hat{\mathbf{E}})\,\hat{\mathbf{E}} = \frac{\mathbf{D}.\mathbf{E}}{E^2}\,\mathbf{E} = \frac{DE\cos\varphi}{E^2}\,\mathbf{E}$$

$$= \frac{k^2}{\mu_0\omega^2}\,\frac{E\cos\varphi}{E}\,\cos\varphi\,\mathbf{E} = \frac{1}{\mu_0}\frac{n^2}{c^2}\frac{n_r^2}{n^2}\,\mathbf{E}$$

$$= \frac{n_r^2}{\mu_0 c^2}\,\mathbf{E}$$

or

$$\mathbf{E} = \frac{\mu_0 c^2}{n_r^2}\,[\mathbf{D}-\mathbf{s}\,(\mathbf{D}.\mathbf{s})] \tag{18.8-21}$$

Using Eqs. (18.8-13) and (18.8-5), we also have

$$\mathbf{D} = \frac{n^2}{c^2\mu_0}\,[\mathbf{E}-\boldsymbol{\kappa}\,(\boldsymbol{\kappa}.\mathbf{E})] \tag{18.8-22}$$

If we write the x, y and z-components of the above equations, then we would obtain

$$\frac{D_j}{\epsilon_j} = \frac{\mu_0 c^2}{n_r^2}\,[D_j - s_j\,(\mathbf{D}.\mathbf{s})] \tag{18.8-23}$$

and

$$\epsilon_j E_j = \frac{n^2}{c^2\mu_0}\,[E_j - \kappa_j\,(\boldsymbol{\kappa}.\mathbf{E})] \tag{18.8-24}$$

where $j = x, y, z$; s_j and κ_j represent the Cartesian components of the unit vectors **s** and $\boldsymbol{\kappa}$, and use has been made of Eq. (18.8-3). Rearranging, we obtain

$$D_j = \frac{s_j\,(\mathbf{D}.\mathbf{s})}{\left(1-\dfrac{n_r^2}{\mu_0 c^2\epsilon_j}\right)} \tag{18.8-25}$$

and

$$E_j = \frac{\kappa_j\,(\boldsymbol{\kappa}.\mathbf{E})}{\left(1-\dfrac{c^2\mu_0\epsilon_j}{n^2}\right)} \tag{18.8-26}$$

Thus

$$\mathbf{D}.\mathbf{s} = \sum_j D_j s_j = \mathbf{D}.\mathbf{s}\sum_j \frac{s_j^2}{\left(1-\dfrac{n_r^2}{\mu_0 c^2\epsilon_j}\right)} \tag{18.8-27}$$

or

$$\sum_j \frac{s_j^2}{\left(1-\dfrac{n_r^2}{\mu_0 c^2\epsilon_j}\right)} = 1 = \sum_j s \tag{18.8-28}$$

The last step follows from the fact that **s** is a unit vector. If we define

$$n_j = c\sqrt{\mu_0 \epsilon_j}, \tag{18.8-29}$$

Eq. (18.8-28) becomes

$$\sum s_j^2 \left[\frac{1}{1 - \left(\frac{n_r}{n_j}\right)^2} - 1 \right] = 0$$

or

$$n_r^2 \sum_j \frac{s_j^2}{n_j^2 - n_r^2} = 0$$

Thus

$$\frac{s_x^2}{n_r^2 - n_x^2} + \frac{s_y^2}{n_r^2 - n_y^2} + \frac{s_z^2}{n_r^2 - n_z^2} = 0 \tag{18.8-30}$$

Similarly

$$\frac{\kappa_x^2}{v_p^2 - v_x^2} + \frac{\kappa_y^2}{v_p^2 - v_y^2} + \frac{\kappa_z^2}{v_p^2 - v_z^2} = 0 \tag{18.8-31}$$

where

$$v_j = \frac{1}{\sqrt{\mu_0 \epsilon_j}};\ j = x, y, z. \tag{18.8-32}$$

Eq. (18.8-30) can be rewritten as

$$s_x^2 (n_r^2 - n_y^2)(n_r^2 - n_z^2) + s_y^2 (n_r^2 - n_x^2)(n_r^2 - n_z^2) + s_z^2 (n_r^2 - n_x^2)(n_r^2 - n_y^2) = 0 \tag{18.8-33}$$

The solution of the above equation will give the variation of n_r with the direction of the ray (i.e. with s_x, s_y and s_z). We label the x, y and z-axes such that

$$n_x > n_y > n_z \tag{18.8-34}$$

We introduce two positive quantities a and b such that

$$n_z^2 = n_y^2 - a \quad \text{and} \quad n_x^2 = n_y^2 + b \tag{18.8-35}$$

and introduce a parameter ξ defined by the following equation:

$$n_r^2 = n_y^2 + \xi \tag{18.8-36}$$

Thus Eq. (18.8-33) reduces to

$$s_x^2 \xi (\xi + a) + s_y^2 (\xi - b)(\xi + a) + s_z^2 (\xi - b)\xi = 0$$

or

$$\xi^2 + (a s_x^2 - b s_y^2 + a s_y^2 - b s_z^2)\xi - ab\, s_y^2 = 0 \tag{18.8-37}$$

where we have used the fact that

$$s_x^2 + s_y^2 + s_z^2 = 1 \tag{18.8-38}$$

If the two roots of Eq. (18.8-37) are denoted by ξ_1 and ξ_2, then

$$\xi_1 \xi_2 = - ab\, s_y^2 \tag{18.8-39}$$

and

$$\xi_1 + \xi_2 = - (as_x^2 - bs_y^2 + as_y^2 - bs_z^2) \tag{18.8-40}$$

Since by definition a and b are positive, ξ_1 and ξ_2 would, in general, be of opposite signs. Thus, if along a certain direction, the two roots are equal (i.e. if $\xi_1 = \xi_2$) then we must have $\xi_1 = \xi_2 = 0$, implying

$$s_y = 0 \text{ and } as_x^2 - bs_z^2 = 0 \tag{18.8-41}$$

The directions along which the ray refractive indices are equal are known as the optic axes. Since $s_y = 0$, these will lie in the x-z plane. If α is the angle made by the optic axes with the z-axis then

$$\tan \alpha = \frac{s_x}{s_z} = \pm \sqrt{\frac{b}{a}} = \pm \sqrt{\frac{n_x^2 - n_y^2}{n_y^2 - n_z^2}} \tag{18.8-42}$$

Thus there will be two ray directions along which the refractive indices will be equal (see Fig. 18.20). If $n_x = n_y$ (i.e. if $\epsilon_x = \epsilon_y$), $\alpha = 0$ (or π) and if $n_z = n_y$ (i.e. if $\epsilon_z = \epsilon_y$) $\alpha = \pi/2$ (or $3\pi/2$); in either case the directions of the two optic axes coincide and hence the medium has only one optic axis; such a medium is known as a uniaxial* medium.

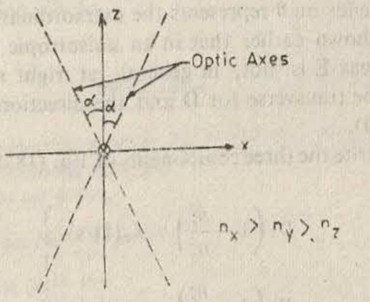


Fig. 18.20 The two directions of the optic axes in a biaxial crystal.

In all what follows we will restrict our attention to uniaxial crystals; we will assume $n_x = n_y$, so that the optic axis is along the z-axis. Thus Eq. (18.8-33) can be written in the form

$$(n_r^2 - n_x^2)\, [s_x^2 (n_r^2 - n_z^2) + s_y^2 (n_r^2 - n_z^2) + s_z^2 (n_r^2 - n_x^2)] = 0 \tag{18.8-43}$$

---

*Crystals in which $n_x = n_y = n_z$ are known as isotropic crystals. Crystals belonging to the cubic system are isotropic. Crystals in which $n_x \neq n_y \neq n_z$ are known as biaxial crystals (because they have two optic axes).

Thus, we must either have

$$n_r = n_x \qquad \text{(ordinary ray)} \quad (18.8\text{-}44)$$

or

$$\begin{aligned} n_r^2 &= n_z^2(1 - s_z^2) + n_x^2 s_z^2 \\ &= n_z^2 \sin^2\theta + n_x^2 \cos^2\theta \quad \text{(extraordinary ray)} \quad (18.8\text{-}45) \end{aligned}$$

where θ is the angle made by the direction of the ray with the optic axis and use has been made of Eq. (18.8-38). Eqs. (18.8-44) and (18.8-45) can be written in the form

$$\mathrm{v}_r = \mathrm{v}_x \qquad \text{(ordinary ray)} \quad (18.8\text{-}46)$$

and

$$\frac{1}{\mathrm{v}_r^2} = \frac{\sin^2\theta}{\mathrm{v}_z^2} + \frac{\cos^2\theta}{\mathrm{v}_x^2} \qquad \text{(extraordinary ray)} \quad (18.8\text{-}47)$$

where we have used Eq. (18.8-20). Eq. (18.8-46) corresponds to the ordinary ray because the ray velocity is independent of the direction of the ray. On the other hand, if we use Eq. (18.8-47) to plot the variation of v_r with θ in the x-z plane we will get an ellipse* as shown in Fig. 18.14. In three-dimensions, Eq. (18.8-47) will represent an ellipsoid of revolution with the optic axis as the axis of revolution. Thus, in a uniaxial medium, for every direction of propagation, the ray can have two velocities, the one which is independent of θ corresponds to the ordinary ray and the other which depends on θ represents the extraordinary ray.

It has been shown earlier that in an anisotropic medium, **D** is at right angle to **k** whereas **E** is not, in general, at right angles to **k**. Thus the wave is said to be transverse for **D** and the direction of vibration is usually shown along **D**.

Now, if we write the three components of Eq. (18.8-25), we would obtain

$$\left.\begin{aligned} D_x\left(1 - \frac{n_r^2}{n_x^2}\right) &= s_x(\mathbf{D}\cdot\mathbf{s}) \\ D_y\left(1 - \frac{n_r^2}{n_x^2}\right) &= s_y(\mathbf{D}\cdot\mathbf{s}) \\ D_z\left(1 - \frac{n_r^2}{n_z^2}\right) &= s_z(\mathbf{D}\cdot\mathbf{s}) \end{aligned}\right\} \qquad (18.8\text{-}48)$$

---

*In the x-z plane, the equation of an ellipse is of the form

$$\frac{x^2}{a^2} + \frac{z^2}{b^2} = 1$$

If we use polar coordinates then $x = r\sin\theta$, $z = r\cos\theta$. Thus the above equation becomes

$$\frac{1}{r^2} = \frac{\sin^2\theta}{a^2} + \frac{\cos^2\theta}{b^2}$$

which is similar to Eq. (18.8-47)

For the ordinary ray, since $n_r = n_x$, we must have

$$\mathbf{D}\cdot\mathbf{s} = 0 \tag{18.8-49}$$

and also $D_z = 0$. Thus **D** is at right angles to the direction of the ray as well as to the optic axis. Further, **D** will be parallel to **E**.

In order to determine the direction of **D** for the extraordinary ray (for which $n_r \neq n_x$), we consider the vector

$$\begin{aligned}\mathbf{N} &= \hat{\mathbf{z}}\times\mathbf{s} = \hat{\mathbf{z}}\times(s_x\,\hat{\mathbf{x}} + s_y\,\hat{\mathbf{y}} + s_z\,\hat{\mathbf{z}})\\ &= s_x\,\hat{\mathbf{y}} - s_y\,\hat{\mathbf{x}}\end{aligned}$$

which will be at right angles to $\hat{\mathbf{z}}$ (which is along the optic axis) and the ray direction **s**. Using Eq. (18.8-48) we get

$$\begin{aligned}\mathbf{D}\cdot\mathbf{N} &= -D_x s_y + D_y s_x = \left[-\frac{n_x^2}{n_x^2 - n_r^2}\,s_x s_y + \frac{n_x^2}{n_x^2 - n_r^2}\,s_x s_y\right]\mathbf{D}.\mathbf{s}\\ &= 0\end{aligned}$$

Thus, the direction of vibration, which is along **D**, lies in the plane containing the optic axis and the ray direction.

The above mathematical analysis justifies the various geometrical constructions made in Sec. 18.5.

PROBLEMS

1. Discuss the state of polarization when the x- and y-components of the electric field are given by the following equations:

   (i) $E_x = E_0 \cos(\omega t + kz)$

   $$E_y = \frac{E_0}{\sqrt{2}} \cos(\omega t + kz + \pi)$$

   (ii) $E_x = E_0 \sin(\omega t + kz)$
   $E_y = E_0 \cos(\omega t + kz)$

   (iii) $E_x = E_0 \sin(kz - \omega t + \pi/3)$
   $E_y = E_0 \sin(kz - \omega t - \pi/6)$

   (iv) $E_z = E_0 \cos(kz - \omega t + \pi/4)$

   $$E_y = \frac{E_0}{\sqrt{2}} \sin(kz - \omega t)$$

2. How will you convert right circularly polarized light into left circularly polarized light?
3. An elliptically polarized light is incident on quartz quarter wave plate. At $z = 0$, the fields are given by (see Fig. 18.18)

   $$E_y = a\cos\omega t$$
   $$E_z = a\cos(\omega t + \pi/3)$$

   What will be the state of polarization of the wave emerging from the crystal?
4. What will be the Brewster angle for a glass slab ($n = 1.5$) immersed in water ($n = 1.33$). (Ans. 48.4°.)

5. Prove that when the angle of incidence corresponds to the Brewster angle, the reflected and refracted rays are at right angles to each other.
6. (a) Consider two crossed polaroids placed in the path of an unpolarized beam of intensity I_0 [see Fig. 18.4 (c)]. If we place a third polaroid in between the two then, in general, some light will be transmitted through. Explain this phenomenon.
   (b) Assuming the pass axis of the third polaroid to be at 45° to the pass axis of either of the polaroids, calculate the intensity of the transmitted beam. Assume that all the polaroids are perfect. (Ans. $1/8\ I_0$).
7. A quarter wave plate is rotated between two crossed polaroids. If an unpolarized beam is incident on the first polaroid, discuss the variation of intensity of the emergent beam as the quarter wave plate is rotated. What would happen if we have a half-wave plate instead of a quarter-wave plate?
8. In the above problem if the optic axis of the quarter-wave plate makes an angle 45° with the pass axis of either polaroids, then show that only a quarter of the incident intensity will be transmitted. If the quarter-wave plate is replaced by a half-wave plate show that half of the incident intensity will be transmitted through.
9. The formulation developed in Sec. 2.4 corresponds to a positive crystal. Modify the analysis for a negative crystal like calcite and show that if the optic axis makes an angle 45° with the surface then for a normally incident beam, the angle of refraction corresponding to the extraordinary ray will be approximately 6°. You may assume $n_0 = 1.658$ and $n_e = 1.486$.
10. For calcite the values of n_0 and n_e for $\lambda = 4046$ Å are 1.68134 and 1.49694 respectively; corresponding to $\lambda = 7065$ Å the values are 1.65207 and 1.48359 respectively. We have a calcite quarter-wave plate corresponding to $\lambda = 4046$ Å. A left circularly polarized beam of $\lambda = 7065$ Å is incident on this plate. Obtain the state of polarization of the emergent beam.
11. When the optic axis lies on the surface of the crystal and in the plane of incidence, show (by geometrical considerations) that the angles of refraction of the ordinary and the extraordinary rays (which we denote by r_0 and r_e respectively) are related through the following equation;

$$\frac{\tan r_0}{\tan r_e} = \frac{n_0}{n_e}$$

    Show that the above result can also be obtained by using Eq. (2.4-10). [In Eq. (2.4-10) we must put $\varphi = \pi/2$, $\theta = r_e - \pi/2$, $n_\parallel = n_0$ and $n_\perp = n_e$].
12. A beam of monochromatic light ($\lambda = 5890$ Å) is incident at an angle of 30° (with the normal to the surface) on a calcite crystal. The optic axis lies on the surface of the crystal and is also in the plane of incidence. Using geometrical method, calculate the angles of refraction for the ordinary and the extraordinary rays and compare the results with the formulae derived in Sec. 2.4.
13. Mica is a biaxial crystal with $n_x = 1.5977$, $n_y = 1.5936$ and $n_z = 1.5601$; these values correspond to the D-lines of sodium. Calculate the angles that the optic axes make with the z-axis. (It may be mentioned that mica cleaves in planes perpendicular to the bisector of the angle between the optic axes).
14. Derive Eq. (18.8-31).
15. Prove that $\epsilon_{xy} = \epsilon_{yx}$, $\epsilon_{yz} = \epsilon_{zy}$ and $\epsilon_{xz} = \epsilon_{zx}$; i.e. the dielectric tensor is symmetric. [Hint: Use the fact that

$$\frac{d}{dt}\left(\tfrac{1}{2}\mathbf{E}.\mathbf{D}\right) = \mathbf{E}.\frac{\partial \mathbf{D}}{\partial t}$$

    The above equality follows from the fact that both sides represent the change of electric energy density; see Eqs. (18.4-3) and (18.4-6).

16. Show that Eq. (18.4–20) represents an ellipse.
[Hint: If the major axis is along the x'-axis, then in the x'–y' coordinate system [see Fig. 18.13 (b)], the equation of an ellipse will be

$$\frac{x'^2}{b_1^2}+\frac{y'^2}{b_2^2}=1$$

But

$$x'=x\cos\varphi+y\sin\varphi \quad \text{and} \quad y'=-x\sin\varphi+y\cos\varphi$$

Thus

$$x^2\left(\frac{\cos^2\varphi}{b_1^2}+\frac{\sin^2\varphi}{b_2^2}\right)+y^2\left(\frac{\sin^2\varphi}{b_1^2}+\frac{\cos^2\varphi}{b_2^2}\right)-xy\sin 2\varphi\left(\frac{1}{b_2^2}-\frac{1}{b_1^2}\right)=1$$

On dividing Eq. (18.4–20) by $\sin^2\theta$ we get

$$\frac{x^2}{a_1^2\sin^2\theta}+\frac{y^2}{a_2^2\sin^2\theta}-2xy\frac{\cos\theta}{a_1a_2\sin^2\theta}=1$$

Comparing the above, two equations we get

$$\frac{1}{a_1^2\sin^2\theta}=\frac{\cos^2\varphi}{b_1^2}+\frac{\sin^2\varphi}{b_2^2}$$

$$\frac{1}{a_2^2\sin^2\theta}=\frac{\sin^2\varphi}{b_1^2}+\frac{\cos^2\varphi}{b_2^2}$$

and

$$\frac{2\cos\theta}{a_1a_2\sin^2\theta}=\sin 2\varphi\left(\frac{1}{b_2^2}-\frac{1}{b_1^2}\right)$$

Simple manipulations will give Eq. (18.4–23)].

REFERENCES AND SUGGESTED READING

ALONSO, M. and FINN, E. J. (1970), *Physics*, Addison Wesley Publishing Co., Inc., Reading, Mass. (USA).

American Institute of Physics (1963), *Polarized Light: Selected Reprints*, American Institute of Physics, New York (USA), 103.

BIRD, G.R. AND PARRISH, M.P. (1960), The Wire grid as a near infrared polarizer, *J. Opt. Soc. Am.*, **50,** 886.

BORN, M. AND WOLF, E. (1970), *Principles of Optics*, Pergamon Press, Oxford (England).

CHANDRASHEKHAR, S. (1956), Simple model for optical activity, *Amer. J. Phys.*, **24,** 503.

FEYNMAN, R.P., LEIGHTON, R.B. AND SANDS, M. (1963), *The Feynman Lectures on Physics*, Vol. I, Addison Wesley Publishing Co., Reading, Mass. (USA).

GAY, P. (1967), *An Introduction to Crystal Optics*, Longmans Green and Co., London (England).

JENKINS, F.A. AND WHITE, H.E. (1976), *Fundamentals of Optics*, McGraw-Hill Book Co., New York (USA).

SHURCLIFF, W.A. AND BALLARD, S.S. (1964), *Polarized Light*, Van Nostrand, Princeton, New Jersey (USA).

WATERMAN, T.H. (1955), Polarized light and animal navigation, *Scien. Amer.*, (July).

WOOD, E.A. (1964), *Crystals and Light*, Van Nostrand Momentum Book No. 5, Van-Nostrand, Princeton, New Jersey (USA).

19 Electromagnetic Waves

And God said "Let there be light"; and there was light.

GENESIS

Nature, and Nature's laws, lay hid in night;
God said, Let Newton be! and all was light.

ALEXANDER POPE

Maxwell could say, when he was finished with his discovery, 'Let there be electricity and magnetism, and there is light.'

FEYNMAN

19.1 Maxwell's Equations

All electromagnetic phenomena can be said to follow from Maxwell's equations. These equations are based on experimental observations and are given by the following equations:

$$\frac{\partial D_x}{\partial x}+\frac{\partial D_y}{\partial y}+\frac{\partial D_z}{\partial z}=\rho \tag{19.1-1}$$

$$\frac{\partial B_x}{\partial x}+\frac{\partial B_y}{\partial y}+\frac{\partial B_z}{\partial z}=0, \tag{19.1-2}$$

$$\left.\begin{aligned}\frac{\partial E_z}{\partial y}-\frac{\partial E_y}{\partial z}&=-\frac{\partial B_x}{\partial t}\\ \frac{\partial E_x}{\partial z}-\frac{\partial E_z}{\partial x}&=-\frac{\partial B_y}{\partial t}\\ \frac{\partial E_y}{\partial x}-\frac{\partial E_x}{\partial y}&=-\frac{\partial B_z}{\partial t}\end{aligned}\right\} \tag{19.1-3}$$

and

$$\left.\begin{aligned}\frac{\partial H_z}{\partial y}-\frac{\partial H_y}{\partial z}&=J_x+\frac{\partial D_x}{\partial t}\\ \frac{\partial H_x}{\partial z}-\frac{\partial H_z}{\partial x}&=J_y+\frac{\partial D_y}{\partial t}\\ \frac{\partial H_y}{\partial x}-\frac{\partial H_x}{\partial y}&=J_z+\frac{\partial D_z}{\partial t}\end{aligned}\right\} \tag{19.1-4}$$

where ρ represents the charge density and **J** the current density; **E, D, B** and **H** represent the electric field, electric displacement, magnetic induction and magnetic field respectively. The physical implications of each equation will be discussed in Sec. 19.9; we may note here that using vector notations the above equations can be written in the following compact form:

$$\text{div } \mathbf{D} = \rho \tag{19.1-5}$$

$$\text{div } \mathbf{B} = 0 \tag{19.1-6}$$

$$\text{curl } \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t} \tag{19.1-7}$$

$$\text{curl } \mathbf{H} = \mathbf{J} + \frac{\partial \mathbf{D}}{\partial t} \tag{19.1-8}$$

The above equations can be solved only if the 'constitutive relations' are known which relate **D** to **E**, **B** to **H** and **J** to **E**. For a linear, isotropic and homogeneous medium the 'constitutive relations' are given by the following equations*

$$\mathbf{D} = \epsilon\, \mathbf{E} \tag{19.1-9}$$

$$\mathbf{B} = \mu\, \mathbf{H} \tag{19.1-10}$$

and

$$\mathbf{J} = \sigma\, \mathbf{E} \tag{19.1-11}$$

where ϵ, μ and σ denote respectively the dielectric permittivity, magnetic permeability and conductivity of the medium respectively.

19.2 Plane Waves in a Dielectric

For a non-charged current-free dielectric, we will have

$$\rho = 0 \tag{19.2-1}$$

and

$$\mathbf{J} = 0 \tag{19.2-2}$$

Further, if we use the 'constitutive relations' [Eqs. (19.1-9)–(19.1-11)], then Eqs. (19.1-1)–(19.1-4) become

$$\frac{\partial E_x}{\partial x} + \frac{\partial E_y}{\partial y} + \frac{\partial E_z}{\partial z} = 0 \tag{19.2-3}$$

$$\frac{\partial H_x}{\partial x} + \frac{\partial H_y}{\partial y} + \frac{\partial H_z}{\partial z} = 0 \tag{19.2-4}$$

---

*It should be mentioned that the 'constitutive relations' depend on the properties of the medium, field strengths, etc. For example, for an anisotropic medium ϵ is a tensor of second rank (see Sec. 18.8); for high field strengths ϵ may itself depend on **E** (see Sec. 15.10); etc.

$$\left.\begin{aligned}\frac{\partial E_z}{\partial y}-\frac{\partial E_y}{\partial z}&=-\mu\frac{\partial H_x}{\partial t}\\ \frac{\partial E_x}{\partial z}-\frac{\partial E_z}{\partial x}&=-\mu\frac{\partial H_y}{\partial t}\\ \frac{\partial E_y}{\partial x}-\frac{\partial E_x}{\partial y}&=-\mu\frac{\partial H_z}{\partial t}\end{aligned}\right\}\qquad(19.2\text{-}5)$$

and

$$\left.\begin{aligned}\frac{\partial H_z}{\partial y}-\frac{\partial H_y}{\partial z}&=\epsilon\frac{\partial E_x}{\partial t}\\ \frac{\partial H_x}{\partial z}-\frac{\partial H_z}{\partial x}&=\epsilon\frac{\partial E_y}{\partial t}\\ \frac{\partial H_y}{\partial x}-\frac{\partial H_x}{\partial y}&=\epsilon\frac{\partial E_z}{\partial t}\end{aligned}\right\}\qquad(19.2\text{-}6)$$

In Sec. 19.3, we will (using the above equations) derive the wave equation; however, in this section we will assume the existence of electromagnetic waves and study the properties of plane waves. For plane waves propagating in the $+z$-direction, the electric and magnetic fields can be written in the form:

$$\mathbf{E}=\mathbf{E}_0\exp\left[i\left(kz-\omega t\right)\right]\qquad(19.2\text{-}7)$$

and

$$\mathbf{H}=\mathbf{H}_0\exp\left[i\left(kz-\omega t\right)\right]\qquad(19.2\text{-}8)$$

where $\mathbf{E}_0$ and $\mathbf{H}_0$ are space and time independent vectors; but may, in general, be complex. Since $\mathbf{E}$ and $\mathbf{H}$ are independent of x and y, Eqs. (19.2-3) and (19.2-4) give

$$\left.\begin{aligned}\frac{\partial E_z}{\partial z}&=0\\ \frac{\partial H_z}{\partial z}&=0\end{aligned}\right\}\qquad(19.2\text{-}9)$$

Since we are interested in plane-wave solutions of the type given by Eqs. (19.2-7) and (19.2-8) we must have*

$$E_z=0\quad\text{and}\quad H_z=0\qquad(19.2\text{-}10)$$

Thus the longitudinal components vanish and hence the waves are of transverse type. Without any loss of generality, we may choose the *x*-axis to be along **E**; thus we may assume

---

*Eq. (19.2-10) does not represent the unique solution of Eq. (19.2-9). For example, we could have chosen $E_z=$ constant; but such a field would not have represented waves.

$$E_y = 0 \tag{19.2-11}$$

Using Eqs. (19.2-7), (19.2-8), (19.2-10) and (19.2-11), Eqs. (19.2-5) and (19.2-6) simplify to

$$0 = i\omega\mu\, H_{0x} \tag{19.2-12a}$$

$$ik\, E_{0x} = i\omega\mu\, H_{0y} \tag{19.2-12b}$$

and

$$-ik\, H_{0y} = -i\omega\epsilon\, E_{0x} \tag{19.2-13a}$$

$$ik\, H_{0x} = 0 \tag{19.2-13b}$$

where

$$\left.\begin{aligned} \mathbf{E}_0 &= \hat{\mathbf{x}}\, E_{0x} + \hat{\mathbf{y}}\, E_{0y} + \hat{\mathbf{z}}\, E_{0z} \\ \mathbf{H}_0 &= \hat{\mathbf{x}}\, H_{0x} + \hat{\mathbf{y}}\, H_{0y} + \hat{\mathbf{z}}\, H_{0z} \end{aligned}\right\} \tag{19.2-14}$$

i.e. E_{0x}, E_{0y} and E_{0z} are the *x*, *y* and *z*-components of $\mathbf{E}_0$, etc. Eqs. (19.2-12) and (19.2-13) give us

$$\left.\begin{aligned} &H_{0x} = 0 \\ \text{and} \qquad &\frac{H_{0y}}{E_{0x}} = \frac{k}{\omega\mu} = \frac{\omega\epsilon}{k} \end{aligned}\right\} \tag{19.2-15}$$

Thus, the plane waves as represented by Eqs. (19.2-7) and (19.2-8) indeed satisfy Maxwell's equations with

$$k^2 = \omega^2\epsilon\mu \tag{19.2-16}$$

The waves are transverse in nature and if we assume the electric vector to be along the *x*-axis then the magnetic vector will be along the *y*-axis so that we may write

$$\left.\begin{aligned} \mathbf{E} &= \hat{\mathbf{x}}\, E_0 e^{i(kz-\omega t)} \\ \mathbf{H} &= \hat{\mathbf{y}}\, H_0 e^{i(kz-\omega t)} \end{aligned}\right\} \tag{19.2-17}$$

with

$$H_0 = \frac{k}{\omega\mu} E_0 \tag{19.2-18}$$

The magnetic induction would, therefore, be given by

$$\mathbf{B} = \hat{\mathbf{y}}\, B_0 e^{i(kz-\omega t)} \tag{19.2-19}$$

where

$$B_0 = \frac{k}{\omega} E_0 = \sqrt{\epsilon\mu}\, E_0 = \frac{1}{\mathrm{v}} E_0 \qquad (19.2\text{-}20a)$$

$\mathrm{v}\left(=\frac{1}{\sqrt{\epsilon\mu}}\right)$ represents the speed of propagation of the electromagnetic waves (see Eq. 19.2-21). In free-space

$$B_0 = \frac{1}{c} E_0 \qquad (19.2\text{-}20b)$$

From the above equations we may draw the following inferences for plane waves:

(i) Both **E** and **H** are at right angles to the direction of propagation. Thus the waves are transverse (see Fig. 19.1).

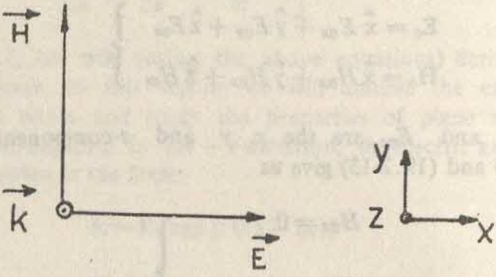


Fig. 19.1 If a plane wave is propagating in the z-direction (which is coming out of the paper) and if at any instant of time the electric vector is along the x-axis then the magnetic vector will be along the y-axis.

(ii) The vectors **E** and **H** are at right angles to each other; thus if the direction of propagation is along the z-axis and if **E** is assumed to point in the x-direction then **H** will point in the y-direction (see Fig. 19.1).

(iii) Since $\frac{k}{\omega\mu}$ is a real number, the electric and magnetic vectors are in phase; thus, if at any instant **E** is zero then **H** is also zero, similarly when **E** attains its maximum value, **H** also attains its maximum value, etc.

(iv) The speed of propagation $\mathrm{v}\left(=\frac{\omega}{k}\right)$ of the wave is given by the following equation:

$$\mathrm{v} = \frac{\omega}{k} = \frac{1}{\sqrt{\epsilon\mu}} \qquad (19.2\text{-}21)$$

For free-space

$$\left.\begin{aligned} \epsilon &= \epsilon_0 = 8.8542 \times 10^{-12} \quad \mathrm{C^2/N\text{-}m^2} \\ \mu &= \mu_0 = 4\pi \times 10^{-7} \quad \mathrm{N\text{-}sec^2/C^2} \end{aligned}\right\} \qquad (19.2\text{-}22)$$

Thus

$$c = \frac{1}{\sqrt{8.8542\times10^{-12}\times4\pi\times10^{-7}}}$$

$$= 2.99794\times10^{8}\ \text{m/sec} \tag{19.2-23}$$

represents the speed of propagation of electromagnetic waves in free space. It may be worthwhile to mention here that Eqs. (19.1-5), (19.1-6) and (19.1-7) were known before Maxwell; he had only introduced the term $\frac{\partial \mathbf{D}}{\partial t}$ in Eq. (19.1-8) (see Sec. 19.9) and it is the presence of this term which leads to the prediction of electromagnetic waves.* Maxwell, using the then known value of ϵ_0 found that the speed of electromagnetic waves should be 3.1074×10^{8} m/sec; this he found to be very close to the velocity of light which was, at that time, known to be 3.14858×10^{8} m/sec (as measured by Fizeau in 1849). Just based on these two numbers and with 'faith in the rationality of nature' he propounded the electromagnetic theory of light and predicted that light must be an electromagnetic wave. In the words of Maxwell himself, the speed of electromagnetic waves

> "....calculated from the electromagnetic measurements of Kohlrausch and Weber, agrees so exactly with the velocity of light calculated from the optical experiments of M. Fizeau, that we can scarcely avoid the inference that *light consists in the transverse undulations of the same medium which is the cause of electric and magnetic phenomena.*"

(v) The refractive index (n) of a dielectric (characterized by dielectric permittivity ϵ and magnetic permeability μ) would be given by

$$n = \frac{c}{v} = \sqrt{\frac{\epsilon\mu}{\epsilon_0\mu_0}} \tag{19.2-24}$$

For most dielectrics $\mu \approx \mu_0$ and one obtains

$$n = \sqrt{\frac{\epsilon}{\epsilon_0}} = \sqrt{\kappa} \tag{19.2-25}$$

where $\kappa\,(=\epsilon/\epsilon_0)$ is known as the dielectric constant of the medium.

(vi) The electric and magnetic waves are interdependent; neither can exist without the other. Physically, an electric field varying in time produces a magnetic field varying in space and time; this changing magnetic field produces an electric field varying in space and time and so

---

*It can be easily seen that in the absence of the term $\epsilon\frac{\partial \mathbf{E}}{\partial t}$ on the RHS of Eq. (19.2-6) the plane waves, as described by Eqs. (19.2-7) and (19.2-8), will *not* represent a solution of Eqs. (19.2-3)—(19.2-6).

on. This mutual generation of electric and magnetic fields result in the propagation of the electromagnetic wave.

(vii) Since Maxwell's equations are linear in **E** and **H** so, if $(\mathbf{E}_1, \mathbf{H}_1)$ and $(\mathbf{E}_2, \mathbf{H}_2)$ are two independent solutions of the Maxwell's equations then $(\mathbf{E}_1 + \mathbf{E}_2, \mathbf{H}_1 + \mathbf{H}_2)$ will also be a solution of the Maxwell's equations. This is the superposition principle according to which the resultant displacement produced by two independent disturbances is the vector sum of the displacements produced by the disturbances independently* (see Chapter 10).

(viii) The plane wave as represented by Eq. (19.2-19) is said to be linearly polarized because the electric vector is always along the x-axis and, similarly the magnetic vector is always along the y-axis (see Fig. 19.2). However, we may consider two independent plane waves**

$$\mathbf{E}_1 = \hat{\mathbf{x}} E_0 \cos (kz - \omega t)$$

$$\mathbf{H}_1 = \hat{\mathbf{y}} H_0 \cos (kz - \omega t)$$

and

$$\mathbf{E}_2 = \hat{\mathbf{y}} E_0 \cos \left(kz - \omega t - \frac{\pi}{2} \right) = \hat{\mathbf{y}} E_0 \sin (kz - \omega t)$$

$$\mathbf{H}_2 = - \hat{\mathbf{x}} H_0 \cos \left(kz - \omega t - \frac{\pi}{2} \right) = - \hat{\mathbf{x}} H_0 \sin (kz - \omega t)$$

then

$$\mathbf{E} = \mathbf{E}_1 + \mathbf{E}_2 = E_0 [\hat{\mathbf{x}} \cos (kz - \omega t) + \hat{\mathbf{y}} \sin (kz - \omega t)]$$

$$\mathbf{H} = \mathbf{H}_1 + \mathbf{H}_2 = H_0 [\hat{\mathbf{y}} \cos (kz - \omega t) - \hat{\mathbf{x}} \sin (kz - \omega t)]$$

will also be a solution which will represent a circularly polarized wave.***

(ix) There exists a wide and continuous variation in the frequency (and wavelength) of electromagnetic waves. The electromagnetic spectrum is shown in Fig. 19.3. The radio waves correspond to wavelengths in the range 10–1000 m whereas the wavelengths of X-rays are in the region of Angstroms ($1 \text{Å} = 10^{-10}$ m). The range of wavelengths of various kinds of electromagnetic waves are shown in Fig. 19.3 and as can be seen, the visible region (4×10^{-5} cm $< \lambda < 7 \times 10^{-5}$ cm) occupies a very small portion of the spectrum. The method of production of different kinds of waves

---

*Thus the superposition principle is a consequence of the linearity of the Maxwell's equations. If, for example, the fields associated with the electromagnetic wave are so high that the dielectric permittivity ϵ depends on **E** itself, then Maxwell's equations will become non-linear and superposition principle will not remain valid. Indeed when we discuss any nonlinear phenomenon, the superposition principle does not hold.

**We are considering the real part of the exponentials appearing on the RHS of Eq. (19.2-17).

***See also Sec. 18.4.

are different; for example, X-rays are usually produced by the sudden stopping or deflection of electrons whereas radio waves may be produced by varying the charge on an antenna. However, all wavelengths propagate with the same speed in vacuum and are always produced by accelerated charges.

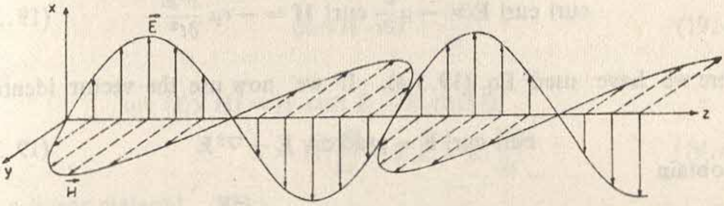


Fig. 19.2 The arrows represent the direction and magnitude of the **E** and **H** vectors (at a particular instant of time) for a plane polarized wave. The electric vectors always lie in the *x-z* plane and the magnetic vectors (shown by dotted lines) lie in the *y-z*-plane.

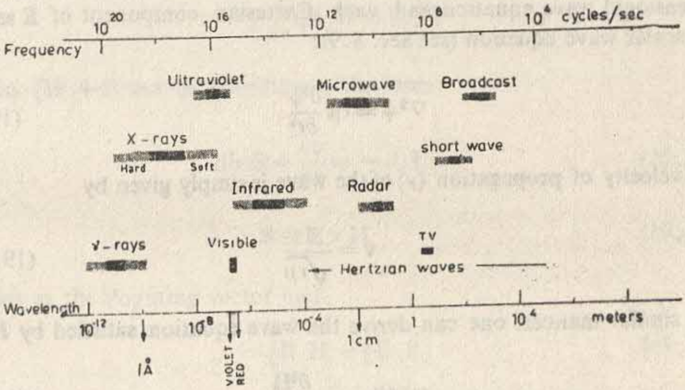


Fig. 19.3 The electromagnetic spectrum.

19.3 The Three-dimensional Wave Equation in a Dielectric

Once again, we consider a non-charged current-free dielectric ($\rho = 0$, $\mathbf{J} = 0$) and for such a case, using the 'constitutive relations' we had obtained Eqs. (19.2-3)—(19.2-6). Using vector notation, these equations can be written in the following compact form:

$$\text{div}\,\mathbf{E} = 0 \qquad (19.3\text{-}1)$$

$$\text{div}\,\mathbf{H} = 0 \qquad (19.3\text{-}2)$$

$$\text{curl}\,\mathbf{E} = -\mu \frac{\partial \mathbf{H}}{\partial t} \qquad (19.3\text{-}3)$$

$$\text{curl}\,\mathbf{H} = \epsilon \frac{\partial \mathbf{E}}{\partial t} \qquad (19.3\text{-}4)$$

In Sec. 19.2 we had shown that plane-wave solutions indeed satisfy the above equations. In this section we will show that the wave equation can be derived from the above equations.

If we take the curl of Eq. (19.3-3), we would obtain

$$\text{curl curl } \mathbf{E} = -\mu\frac{\partial}{\partial t}\text{ curl } \mathbf{H} = -\epsilon\mu\frac{\partial^2 \mathbf{E}}{\partial t^2} \qquad (19.3\text{-}5)$$

where we have used Eq. (19.3-4). If we now use the vector identity*

$$\text{curl curl } \mathbf{E} = \text{grad div } \mathbf{E} - \nabla^2 \mathbf{E} \qquad (19.3\text{-}6)$$

we obtain

$$\text{grad div } \mathbf{E} - \nabla^2 \mathbf{E} = -\epsilon\mu\frac{\partial^2 \mathbf{E}}{\partial t^2} \qquad (19.3\text{-}7)$$

or

$$\nabla^2 \mathbf{E} = \epsilon\mu\frac{\partial^2 \mathbf{E}}{\partial t^2} \qquad (19.3\text{-}8)$$

where we have used Eq. (19.3-1). Eq. (19.3-8) is known as the three-dimensional wave equation and each Cartesian component of **E** satisfies the scalar wave equation (see Sec. 8.9):

$$\nabla^2 \psi = \epsilon\mu\frac{\partial^2 \psi}{\partial t^2} \qquad (19.3\text{-}9)$$

The velocity of propagation (v) of the wave is simply given by

$$v = \frac{1}{\sqrt{\epsilon\mu}} \qquad (19.3\text{-}10)$$

In a similar manner, one can derive the wave equation satisfied by **H**

$$\nabla^2 \mathbf{H} = \epsilon\mu\frac{\partial^2 \mathbf{H}}{\partial t^2} \qquad (19.3\text{-}11)$$

It can be easily seen that the solutions expressed by Eqs. (19.2-7) and (19.2-8) indeed satisfy Eqs. (19.3-8) and (19.3-11) provided

$$\frac{\omega^2}{k^2} = \frac{1}{\epsilon\mu} \qquad (19.3\text{-}12)$$

---

*It may be worthwhile to point out that the operator ∇^2 is *defined* by the following equation:

$$\nabla^2 \mathbf{E} \equiv \text{grad div } \mathbf{E} - \text{curl curl } \mathbf{E}$$

However,

$$(\nabla^2 \mathbf{E})_x = \text{div grad } E_x = \frac{\partial^2 E_x}{\partial x^2} + \frac{\partial^2 E_x}{\partial y^2} + \frac{\partial^2 E_x}{\partial z^2}$$

i.e. a cartesian component of $\nabla^2\mathbf{E}$ is the div grad of the Cartesian component and

$$(\nabla^2 E)_r \neq \text{div grad } E_r$$

19.4 The Poynting Vector

Let us consider Eqs. (19.1-7) and (19.1-8)

$$\text{curl}\,\mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t} \qquad (19.4\text{-}1)$$

$$\text{curl}\,\mathbf{H} = \mathbf{J} + \frac{\partial \mathbf{D}}{\partial t} \qquad (19.4\text{-}2)$$

Now

$$\begin{aligned} \text{div}\,(\mathbf{E}\times\mathbf{H}) &= \mathbf{H}.\text{curl}\,\mathbf{E} - \mathbf{E}.\text{curl}\,\mathbf{H} \\ &= -\mathbf{H}.\frac{\partial \mathbf{B}}{\partial t} - \mathbf{J}.\mathbf{E} - \mathbf{E}.\frac{\partial \mathbf{D}}{\partial t} \qquad (19.4\text{-}3) \end{aligned}$$

For a linear material

$$\begin{aligned} \mathbf{H}.\frac{\partial \mathbf{B}}{\partial t} + \mathbf{E}.\frac{\partial \mathbf{D}}{\partial t} &= \mu\mathbf{H}.\frac{\partial \mathbf{H}}{\partial t} + \epsilon\mathbf{E}.\frac{\partial \mathbf{E}}{\partial t} \\ &= \tfrac{1}{2}\mu\frac{\partial}{\partial t}(\mathbf{H}.\mathbf{H}) + \tfrac{1}{2}\epsilon\frac{\partial}{\partial t}(\mathbf{E}.\mathbf{E}) \\ &= \tfrac{1}{2}\frac{\partial}{\partial t}(\mathbf{B}.\mathbf{H} + \mathbf{D}.\mathbf{E}) \end{aligned}$$

Thus Eq. (19.4-3) can be rewritten in the form

$$\text{div}\,\mathbf{S} + \frac{\partial u}{\partial t} = -\mathbf{J}.\mathbf{E} \qquad (19.4\text{-}4)$$

where

$$\mathbf{S} \equiv \mathbf{E}\times\mathbf{H} \qquad (19.4\text{-}5)$$

is known as the Poynting vector and

$$u = \tfrac{1}{2}\mathbf{B}.\mathbf{H} + \tfrac{1}{2}\mathbf{D}.\mathbf{E} \qquad (19.4\text{-}6)$$

Eq. (19.4-4) resembles the equation of continuity and for a physical interpretation we note that if a charge q (moving with velocity $\mathbf{v}$) is acted on by an electromagnetic field then the work done by the field in moving it through a distance $d\mathbf{s}$ would be $\mathbf{F}.d\mathbf{s}$; thus the work done per unit time would be

$$\begin{aligned} \mathbf{F}.\frac{d\mathbf{s}}{dt} &= \mathbf{F}.\mathbf{v} \\ &= [q\mathbf{E} + q\mathbf{v}\times\mathbf{B}].\mathbf{v} \\ &= q\mathbf{E}.\mathbf{v} \qquad (19.4\text{-}7) \end{aligned}$$

If there are N charged particles per unit volume, each carrying a charge q then the work done per unit volume would be

$$Nq\mathbf{v}.\mathbf{E} = \mathbf{J}.\mathbf{E} \qquad (19.4\text{-}8)$$

where **J** represents the current density. The energy appears in the form of kinetic (or heat) energy of the charged particles. Thus the term **J**.**E** represents the familiar Joule loss and, therefore, the quantity **J**.**E** appearing on the R.H.S. of Eq. (19.4-4) would represent the rate at which energy is produced per unit volume per unit time. Consequently, we may interpret Eq. (19.4-4) as an equation of continuity* for energy with u representing the energy per unit volume. The quantities $\frac{1}{2}\mathbf{D}.\mathbf{E}$ and $\frac{1}{2}\mathbf{B}.\mathbf{H}$ represent the electrical and magnetic energies per unit volume respectively. Further, we may interpret **S**.d**a** as the electromagnetic energy crossing the area d**a** per unit time. For plane waves in a dielectric, we may write

$$\left.\begin{aligned} &\mathbf{E} = \hat{\mathbf{x}}\,\mathrm{E}_0 \cos(kz - \omega t) \\ \text{and}\quad &\mathbf{H} = \hat{\mathbf{y}}\,\mathrm{H}_0 \cos(kz - \omega t) = \hat{\mathbf{y}}\,\frac{k}{\omega\mu}\,\mathrm{E}_0 \cos(kz - \omega t) \end{aligned}\right\} \qquad (19.4\text{-}9)$$

Thus

$$\begin{aligned} \mathbf{S} &= \mathbf{E}\times\mathbf{H} \\ &= \hat{\mathbf{z}}\,\frac{k}{\omega\mu}\,E_0^2 \cos^2(kz - \omega t) \end{aligned} \qquad (19.4\text{-}10)$$

which implies that the energy flow is in the z-direction (which represents the direction of propagation of the wave) and that an amount of energy

$$\frac{k}{\omega\mu}\,E_0^2 \cos^2(kz - \omega t) \qquad (19\ 4\text{-}11)$$

crosses a unit area (perpendicular to the z-axis) per unit time. For optical beams $\omega \approx 10^{15}\ \mathrm{sec}^{-1}$ and the $\cos^2$ term fluctuates with extreme rapidity** and any detector would record only an average value. Since

$$\begin{aligned} \langle \cos^2(kz - \omega t) \rangle &= \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{+T} \cos^2(kz - \omega t)\, dt \\ &= \tfrac{1}{2}, \end{aligned} \qquad (19.4\text{-}12)$$

we obtain

$$\langle \mathbf{S} \rangle = \hat{\mathbf{z}}\,\frac{k}{2\omega\mu}\,E_0^2 \qquad (19.4\text{-}13)$$

where $\langle \ldots \rangle$ denotes the time-average value of the quantity inside the angular brackets.

---

*The equation of continuity is always written in the form

$$\operatorname{div} \mathbf{J} + \frac{\partial\rho}{\partial t} = 0$$

where ρ represents the charge density and **J** the current density; i.e. **J**.d**a** represents the amount of charge crossing the area d**a** per unit time.

**See also Sec. 11.3.

We must hasten to point out that $\mathbf{S}.d\mathbf{a}$ does not *always* represent the rate of energy flow through the area $d\mathbf{a}$; for example, we may have *static* electric and magnetic fields where $\mathbf{E}\times\mathbf{H}$ is finite but we know that there is no energy flow. However, the integral

$$\oint \mathbf{S}.d\mathbf{a} \tag{19.4-14}$$

over a closed surface rigorously represents the net energy flowing out of the surface. This follows immediately if we carry out a volume integral of Eq. (19.4-4) to give

$$\int \operatorname{div}\mathbf{S}\,dV + \frac{\partial}{\partial t}\int u\,dV = -\int \mathbf{J}\cdot\mathbf{E}\,dV \tag{19.4-15}$$

or

$$-\frac{\partial}{\partial t}\int u\,dV = \oint \mathbf{S}\cdot d\mathbf{a} + \int \mathbf{J}\cdot\mathbf{E}\,dV \tag{19.4-16}$$

where we have used the divergence theorem. The quantity on the LHS represents the rate of decrease of the total energy, this must be equal to the Joule loss plus the net flow out of the surface enclosing the volume.

19.4.1 The Oscillating Dipole

Consider an oscillating dipole in the z-direction:

$$\mathbf{p} = p_0 e^{-i\omega t}\hat{\mathbf{z}}$$

At large distances from such a dipole the fields are of the form* (see Fig. 19.4)

$$\mathbf{E} = -\left(\frac{k^2 p_0}{4\pi\epsilon_0}\right)\sin\theta\,\frac{e^{i(kr-\omega t)}}{r}\,\hat{\theta} \tag{19.4-17}$$

$$\mathbf{H} = -\left(\frac{\omega k p_0}{4\pi}\right)\sin\theta\,\frac{e^{i(kr-\omega t)}}{r}\,\hat{\varphi} \tag{19.4-18}$$

where $k = \omega\sqrt{\epsilon_0\mu_0}$ and other symbols have their usual meaning. Notice that the fields fall off as $1/r$ and that they are in phase. Further, the ratio of the amplitudes of the magnetic and electric fields is

$$\frac{\omega k p_0/4\pi}{k^2 p_0/4\pi\epsilon_0} = \frac{\omega\epsilon_0}{k} = \frac{k}{\omega\mu_0} \tag{19.4-19}$$

which is consistent with Eq. (19.2-20).

---

*See, for example, Panofsky and Phillips (1962) p. 258.

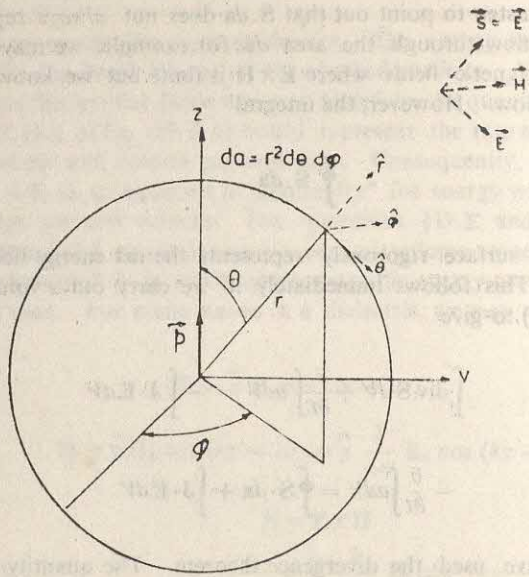


Fig. 19.4 The direction of the electric and magnetic fields and of the Poynting vector from an oscillating dipole. In order to calculate the total energy radiated per unit time, we must integrate the Poynting vector over the surface of a sphere.

Thus*

$$\mathbf{S} = \mathbf{E} \times \mathbf{H}$$
$$= \frac{\omega k^3 p_0^2}{16\pi^2 \epsilon_0} \sin^2\theta \, \frac{\cos^2(kr - \omega t)}{r^2} \hat{\mathbf{r}} \tag{19.4-20}$$

Eq. (19.4-20) shows that **S** falls off as $1/r^2$ as it indeed should be for a spherical wave (this is the inverse square law). If we integrate over a sphere of radius r, we would obtain

$$P = \oint \mathbf{S} \cdot d\mathbf{a} = r^2 \iint \mathbf{S} \cdot \hat{\mathbf{r}} \sin\theta \, d\theta \, d\varphi$$
$$= \frac{\omega k^3 p_0^2}{16\pi^2 \epsilon_0} \cos^2(kr - \omega t) \int_0^{\pi} \sin^2\theta \sin\theta \, d\theta \int_0^{2\pi} d\varphi$$
$$= \frac{\omega k^3 p_0^2}{6\pi \epsilon_0} \cos^2(kr - \omega t) \tag{19.4-21}$$

---

*In order to calculate the Poynting vector we must take the products of the real parts of **E** and **H**. We may note that in the complex representation, if $\mathbf{E} = \mathbf{E}_1 + \mathbf{E}_2$ then

$$\mathrm{Re}(\mathbf{E}) = \mathrm{Re}(\mathbf{E}_1) + \mathrm{Re}(\mathbf{E}_2)$$

However,

$$(\mathrm{Re}\,\mathbf{E}_1) \times (\mathrm{Re}\,\mathbf{E}_2) \neq \mathrm{Re}(\mathbf{E}_1 \times \mathbf{E}_2)$$

Here $\mathrm{Re}(\mathbf{E})$ denotes the real part of **E**.

where P represents the instantaneous radiated power. Since the $\cos^2$ term fluctuates very rapidly, the average radiated power would be given by

$$\bar{P} = \frac{\omega k^3 p_0^2}{12\pi\epsilon_0} \tag{19.4-22}$$

19.5 Energy Density and Intensity of an Electromagnetic Wave

In Sec. 19.4 we have shown that the energy/unit volume associated with a plane wave is given by

$$u = \tfrac{1}{2}\mathbf{D}\cdot\mathbf{E} + \tfrac{1}{2}\mathbf{B}\cdot\mathbf{H} = \tfrac{1}{2}\epsilon E^2 + \frac{1}{2\mu} B^2 \tag{19.5-1}$$

For a linearly polarized plane wave, we may write

$$E_x = E_0 \cos(kz - \omega t),\ E_y = 0,\ E_z = 0 \tag{19.5-2}$$

$$B_x = 0,\ B_y = B_0 \cos(kz - \omega t),\ B_z = 0 \tag{19.5-3}$$

Thus

$$u = \tfrac{1}{2}\epsilon E_0^2 \cos^2(kz - \omega t) + \frac{1}{2\mu} B_0^2 \cos^2(kz - \omega t)$$

Since $B_0 = \sqrt{\epsilon\mu}\, E_0$ (see Eq. 19.2-20a), we get

$$\frac{B_0^2}{2\mu} = \tfrac{1}{2}\epsilon E_0^2 \tag{19.5-4}$$

Thus the energy associated with the electric field is equal to the energy associated with the magnetic field. If we take the time average of the $\cos^2$ terms, we would get

$$< u > = \tfrac{1}{2}\epsilon E_0^2 \tag{19.5-5}$$

Further, in order to obtain the intensity of the beam we must multiply $< u >$ by the speed of propagation which will give us the energy crossing an unit area in unit time. Thus, the intensity is given by

$$I = \tfrac{1}{2}\epsilon v E_0^2 = \tfrac{1}{2}\sqrt{\epsilon/\mu}\, E_0^2 \tag{19.5-6}$$

This should be consistent with Eq. (19.4-13). Indeed, if we substitute $k = \omega\sqrt{\epsilon\mu}$ in Eq. (19.4-13), we would obtain Eq. (19.5-6). In free space

$$\begin{aligned} I &= \tfrac{1}{2}\epsilon_0 c E_0^2 \\ &= \tfrac{1}{2}(8.854\times 10^{-12}\,\text{coul}^2/\text{N-m}^2)\times(3\times 10^8\,\text{m/sec})\, E_0^2 \\ &= (1.33\times 10^{-3}\,\text{W/V}^2)\, E_0^2 \end{aligned}$$

For example, for a 100 W lamp, the intensity at a distance of 10 m would be

$$I = \frac{100}{4\pi(10)^2} = 7.8 \times 10^{-2}\ \text{W/m}^2$$

where we have assumed light to spread out uniformly in all directions. Thus

$$E_0 = \left[\frac{7.8 \times 10^{-2}}{1.33 \times 10^{-3}}\right]^{1/2} \approx 7.66\ \text{V/m}$$

It is of interest to mention here that since a laser beam is almost perfectly parallel, it can be focussed by a lens to a cross-sectional area of less than 10^{-6} sq cm. Thus for a 10^5 watt laser beam, the intensity at the focal plane would be

$$I = \frac{10^5}{10^{-10}} = 10^{15}\ \text{W/m}^2$$

Thus

$$E_0 = \left[\frac{10^{15}}{1.33 \times 10^{-3}}\right]^{1/2} \approx 0.75 \times 10^9\ \text{V/m}.$$

Such high electric fields can cause extreme high temperatures which may result in the burning of a target.

19.6. Radiation Pressure*

Let us consider a linearly polarized electromagnetic wave propagating in the $+z$-direction; we assume the electric field to be along the x-direction and the magnetic field along the y-direction (see Fig. 19.1). The electromagnetic wave is assumed to interact with a charge q; the electric field makes the charge move up and down along the x-axis. Thus the charge acquires a certain velocity in the x-direction and since the magnetic field is along the y-axis, a force

$$\mathbf{F} = q\mathbf{v} \times \mathbf{B} \tag{19.6-1}$$

acts on the charge q. This force acts along the z-axis** (i.e. along the

---

*Reference: Feynman *et al.* (1963) Sec. 34-9. A rigorous analysis has been given by Panofsky & Phillips (1962) Chapter 10.

**In Sec. 6.2, we had shown that in the presence of a field $\mathbf{E} = \hat{\mathbf{x}} E_0 \cos(kz - \omega t)$, the displacement is given by (see Eq. 6.2-3):

$$\mathbf{x} = \hat{\mathbf{x}}\, q\, E_0 A \cos(kz - \omega t + \varphi)$$

where we have explicitly shown that the amplitude is proportional to q and E_0. Thus

$$\mathbf{v} = \dot{\mathbf{x}} = +\hat{\mathbf{x}} q E_0 A\omega \sin(kz - \omega t + \varphi)$$

Now $\quad \mathbf{B} = \hat{\mathbf{y}} B_0 \cos(kz - \omega t) = \hat{\mathbf{y}} \frac{E_0}{c} \cos(kz - \omega t)$. Thus

(*Contd.*)

direction of propagation of the wave) and constitutes what is known as 'radiation pressure'. Thus,

$$\mathbf{F} = q\mathrm{v}\, B\hat{\mathbf{z}} \tag{19.6-2}$$

But

$$B = \frac{E}{c} \tag{19.6-3}$$

(see Eq. 19.2-20). Thus

$$\mathbf{F} = \frac{qE\mathrm{v}}{c}\,\hat{\mathbf{z}} \tag{19.6-4}$$

Now $qE\mathrm{v}$ represents the work done by the field on the charge per unit time (see Sec. 19.4); thus, if we consider an unit volume then

$$\mathbf{F} = \frac{1}{c}\frac{du}{dt}\,\hat{\mathbf{z}} \tag{19.6-5}$$

But the force is equal to the change in momentum per unit time. Consequently, the momentum per unit volume associated with the plane wave would be given by

$$\mathbf{p} = \frac{u}{c}\,\hat{\mathbf{z}} \tag{19.6-6}$$

In Chapter 21, we will show that light essentially consists of corpuscles called photons. Each photon carries an energy equal to $h\nu$; the photon momentum would, therefore, be given by

$$p = \frac{h\nu}{c} \tag{19.6-7}$$

Let us consider a plane wave incident normally on a perfect absorber If we consider an area dS on the absorbing surface then the momentum transferred to the area dS in time cdt would be

$$p\,dS\,cdt$$

which represents the momentum contained in a cylindrical volume $dS\,cdt$ (see Fig. 19.5). Thus the force acting on the area dS would be

$$pc\,dS$$

Hence

$$P_{\text{rad}} = cp = u \tag{19.6-8}$$

$$\mathbf{F} = q\mathbf{v} \times \mathbf{B} = +\hat{\mathbf{z}}q^2\,\frac{E_0^2\omega}{c}\,A\,[\cos(kz-\omega t)]$$
$$[\sin(kz-\omega t)\cos\varphi + \cos(kz-\omega t)\sin\varphi]$$

If we carry out a time averaging, then

$$<\mathbf{F}> = \hat{\mathbf{z}}\,\frac{q^2E_0^2\omega}{2c}\,A\sin\varphi$$

Since $\sin\varphi$ is always positive (see Eq. 6.2-7) the force is always in the z-direction.

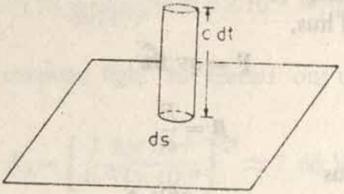


Fig. 19.5 A cylindrical volume to calculate radiation pressure.

where P_{rad} represents the radiation pressure due to a plane wave incident on a perfect absorber. On the other hand, for a perfect reflector, the momentum of the reflected wave is equal and opposite of the momentum associated with the incident wave. Thus the momentum transferred would be twice of the above value and hence

$$P_{\text{rad}} = 2cp = 2u \tag{19.6-9}$$

In order to have a numerical appreciation, let us consider a light beam of intensity $I = 3000 \text{W/m}^2$ falling on a perfectly reflecting mirror. Since $I = cu$, we have

$$u = \frac{3000 \text{ W/m}^2}{3 \times 10^8 \text{ m/sec}} = 10^{-6} \text{ J/m}^3$$

The radiation pressure would be

$$10^{-6} \text{N/m}^2$$

which may be compared with the atmospheric pressure ($\simeq 10^5 \text{N/m}^2$).

It has been possible to measure the radiation pressure by allowing a light beam to fall on a highly polished mirror *M* (see Fig. 19.6).* The

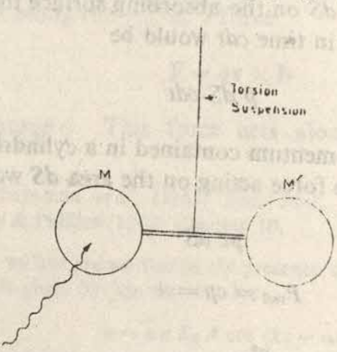


Fig. 19.6 An experimental arrangement to measure radiation pressure.

---

*The experiment was first carried out by Lebedev in Russia in 1899; experimental arrangement shown in Fig. 19.6 is of Nichols and Hull who performed the experiment in 1901 and confirmed the prediction of radiation pressure.

radiation pressure caused a twist in the suspension which was measured. The intensity of the beam can be determined by allowing it to fall on an absorber (like a blackened disc) and measuring the temperature rise. In a particular run, the radiation pressure was found to be about $7.01 \times 10^{-6}\,\text{N/m}^2$ which was in good agreement with the predicted value of $7.05 \times 10^{-6}\,\text{N/m}^2$.

For oblique incidence on a perfect reflector, the change in momentum per unit volume would be $2p \cos \theta$ and the radiation pressure would be

$$P_{\text{rad}} = 2cp \cos^2 \theta = 2u \cos^2 \theta \tag{19.6-10}$$

where θ represents the angle of incidence.

19.7 The Wave Equation in a Conducting Medium

In Sec. 19.3 we had assumed $\mathbf{J} = 0$. For a conducting medium

$$\mathbf{J} = \sigma \mathbf{E} \tag{19.7-1}$$

where σ represents the conductivity of the medium. Thus, Maxwell's equations become

$$\text{div}\, \mathbf{E} = 0 \tag{19.7-2}$$

$$\text{div}\, \mathbf{H} = 0 \tag{19.7-3}$$

$$\text{curl}\, \mathbf{E} = -\mu \frac{\partial \mathbf{H}}{\partial t} \tag{19.7-4}$$

$$\text{curl}\, \mathbf{H} = \sigma \mathbf{E} + \epsilon \frac{\partial \mathbf{E}}{\partial t} \tag{19.7-5}$$

Taking the curl of Eq. (19.7-4), we get

$$\text{curl curl}\, \mathbf{E} = -\mu \frac{\partial}{\partial t} \text{curl}\, \mathbf{H}$$

or

$$\text{grad div}\, \mathbf{E} - \nabla^2 \mathbf{E} = -\mu\sigma \frac{\partial \mathbf{E}}{\partial t} - \mu\epsilon \frac{\partial^2 \mathbf{E}}{\partial t^2}$$

Using Eq. (19.7-2), we get

$$\nabla^2 \mathbf{E} - \mu\sigma \frac{\partial \mathbf{E}}{\partial t} - \mu\epsilon \frac{\partial^2 \mathbf{E}}{\partial t^2} = 0, \tag{19.7-6}$$

which is the wave equation for a conducting medium. For a plane wave of the type

$$\mathbf{E} = \mathbf{E}_0 \exp [i(kz - \omega t)] \tag{19.7-7}$$

we obtain

$$-k^2 + i\omega\mu\sigma + \omega^2 \epsilon\mu = 0 \tag{19.7-8}$$

which shows that k must be a complex number. If we write

$$k = \alpha + i\beta \tag{19.7-9}$$

then

$$-(\alpha^2 + 2i\alpha\beta - \beta^2) + i\omega\mu\sigma + \omega^2\epsilon\mu = 0$$

Equating real and imaginary parts, we get

$$\alpha^2 - \beta^2 = \omega^2\epsilon\mu \tag{19.7-10}$$

and

$$\beta = \frac{\omega\mu\sigma}{2\alpha} \tag{19.7-11}$$

Substituting for β in Eq. (19.7-10) and solving for α, we get

$$\alpha = \omega\sqrt{\epsilon\mu}\left[\frac{1}{2} \pm \sqrt{1 + \frac{\sigma^2}{\omega^2\epsilon^2}}\right]^{1/2} \tag{19.7-12}$$

We must choose the positive sign, the negative sign would make α complex. Thus

$$\left.\begin{aligned} \alpha &= \omega\sqrt{\epsilon\mu}\left[\frac{1}{2} + \frac{1}{2}\left(1 + \frac{\sigma^2}{\omega^2\epsilon^2}\right)^{1/2}\right]^{1/2} \\ \text{and} \qquad & \\ \beta &= \frac{\omega\mu\sigma}{2\alpha} \end{aligned}\right\} \tag{19.7-13}$$

Now, when k is complex, Eq. (19.7-7) becomes

$$\mathbf{E} = \mathbf{E}_0 \exp(-\beta z)\exp[i(\alpha z - \omega t)] \tag{19.7-14}$$

which represents an attenuated wave. The attenuation is due to the Joule-loss. For a good conductor

$$\frac{\sigma}{\omega\epsilon} >> 1 \tag{19.7-15}$$

and one obtains

$$\alpha \approx \beta \approx \left(\frac{\omega\mu\sigma}{2}\right)^{1/2} \tag{19.7-16}$$

Indeed if $\frac{\sigma}{\omega\epsilon} << 1$ (say $\gtrsim 0.01$), the medium can be classified as a dielectric and if $\frac{\sigma}{\omega\epsilon} >> 1$ (say $\gtrsim 100$), the medium can be classified as a conductor. For

$$0.01 \lesssim \frac{\sigma}{\omega\epsilon} \lesssim 100,$$

the medium is said to be a quasi-conductor. Thus, depending on the frequency, a particular material can behave as a dielectric or as a conductor. For example, for fresh water $\frac{\epsilon}{\epsilon_0} \approx 80$ and $\sigma \approx 10^{-3}$ mhos/m. (Both ϵ and σ can be assumed to be constants at low frequencies). Thus

$$\frac{\sigma}{\epsilon} \approx \frac{10^{-3}}{80 \times 8.9 \times 10^{-12}} \approx 1.3 \times 10^{6} \text{ sec}$$

For

$$\omega = 2\pi \times 10 \text{ sec}^{-1}$$

$$\frac{\sigma}{\omega\epsilon} \approx 2 \times 10^{4}$$

and for

$$\omega = 2\pi \times 10^{10} \text{ sec}^{-1}$$

$$\frac{\sigma}{\omega\epsilon} \approx 2 \times 10^{-6}$$

Thus, fresh water behaves as a good conductor for $\nu \lesssim 10^{3}$ sec^{-1} and as a dielectric for $\nu \gtrsim 10^{7}$ sec^{-1}. On the other hand, for copper one may assume $\epsilon \approx \epsilon_0$ and $\sigma \approx 5.8 \times 10^{7}$ mhos/m and for $\omega \approx 2\pi \times 10^{10}$ sec^{-1}

$$\frac{\sigma}{\omega\epsilon} \approx \frac{5.8 \times 10^{7}}{2\pi \times 10^{10} \times 8.9 \times 10^{-12}} \approx 10^{8}$$

Thus even for such frequencies it behaves as an excellent conductor.

From Eq. (19.7-14), it can be easily seen that the field decreases by a factor e in traversing a distance

$$\delta = \frac{1}{\beta},$$

which is known as the penetration depth. For copper,

$$\mu \approx \mu_0 = 4\pi \times 10^{-7} \text{ N/amp}^2$$

and

$$\delta \approx \left(\frac{2}{\omega\mu\sigma}\right)^{1/2} \approx \left(\frac{2}{2\pi\nu \times 4\pi \times 10^{-7} \times 5.8 \times 10^{7}}\right)^{1/2}$$

$$\approx 0.065/\sqrt{\nu} \text{ m}$$

Thus for $\nu \approx 100$ sec^{-1}, $\delta \approx 0.0065 \text{ m} = 0.65$ cm whereas for $\nu \approx 10^{8}$ sec^{-1}, $\delta \approx 6.5 \times 10^{-6}$ m showing that the penetration decreases with increase in frequency.

19.8 The Continuity Conditions

In this section we will derive the continuity conditions for electric and magnetic fields at the interface of two media. Let us first consider the equation

$$\text{div}\, \mathbf{B} = 0 \tag{19.8-1}$$

At the interface of two media, we consider a pill box which encloses an area $\triangle S$ of the interface (see Fig. 19.7). Let the height of the pill

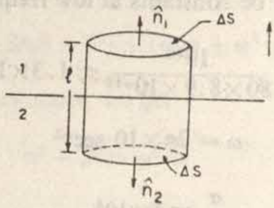


Fig. 19.7. A cylindrical pill box at the interface of two dielectrics.

box be l. Now if we integrate div **B** over the cylindrical volume then, using Gauss' theorem we obtain

$$0=\int \operatorname{div}\mathbf{B}\,dV=\oint_{S_1}\mathbf{B}\cdot dV+\oint_{S_2}\mathbf{B}\cdot d\mathbf{a}+\oint_{S_3}\mathbf{B}\cdot d\mathbf{a}$$

where S_3 represents the curved surface of the cylinder. If we let $l\to 0$ then the third integral vanishes and we obtain

$$\oint_{S_1}\mathbf{B}\cdot d\mathbf{a}=-\oint_{S_2}\mathbf{B}\cdot d\mathbf{a}$$

or

$$\mathbf{B}_1\cdot\hat{\mathbf{n}}_1\,\triangle S=-\,\mathbf{B}_2\cdot\hat{\mathbf{n}}_2\,\triangle S \tag{19.8-2}$$

or

$$E_{1n}=B_{2n} \tag{19.8-3}$$

where the directions of $\hat{\mathbf{n}}$, $\hat{\mathbf{n}}_1$ and $\hat{\mathbf{n}}_2$ are shown in Fig. 19.7. Thus, the normal component of **B** is continuous across the interface.

Similarly, in the absence of free charges

$$\operatorname{div}\mathbf{D}=0$$

and one obtains*

$$D_{1n}=D_{2n} \tag{19.8-4}$$

showing that the normal component of **D** is also continuous across the interface.

We next consider the equation

$$\operatorname{curl}\mathbf{E}+\frac{\partial\mathbf{B}}{\partial t}=0$$

---

*Rigorously

$$D_{1n}-D_{2n}=\sigma$$

where σ represents the surface charge density.

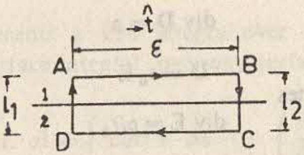


Fig. 19.8 A rectangular loop at the interface of two dielectrics.

We consider a rectangular loop ABCD as shown in Fig. 19.8. Now

$$0 = \oint_S \text{curl}\,\mathbf{E}\cdot d\mathbf{a} + \oint_S \frac{\partial \mathbf{B}}{\partial t}\cdot d\mathbf{a} \qquad (19.8\text{-}5)$$

where the surface integral is over any surface bounding the loop ABCD. Using Stoke's theorem we get

$$\oint \mathbf{E}\cdot d\mathbf{l} = -\int \frac{\partial \mathbf{B}}{\partial t}\cdot d\mathbf{a} \qquad (19.8\text{-}6)$$

or

$$\left(\int_{AB} + \int_{BC} + \int_{CD} + \int_{DA}\right)\mathbf{E}\cdot d\mathbf{l} = -\int \frac{\partial \mathbf{B}}{\partial t}\cdot d\mathbf{a}$$

If we let l_1 and $l_2 \rightarrow 0$ then the integrals along BC and DA tend to zero and since the area of the loop also tends to zero the RHS also vanishes. Thus we obtain

$$\int_{AB} \mathbf{E}\cdot d\mathbf{l} + \int_{BA} \mathbf{E}\cdot d\mathbf{l} = 0$$

or

$$(\mathbf{E}_1\cdot\hat{\mathbf{t}})\,\epsilon + [\mathbf{E}_2\cdot(-\hat{\mathbf{t}})]\,\epsilon = 0$$

or

$$\mathbf{E}_{1t} = \mathbf{E}_{2t}$$

where E_{1t} and E_{2t} represent the tangential components of E which are continuous across the interface.

Similarly, Eq. (19.1-8) gives us*

$$H_{1t} = H_{2t}$$

In summary, in the absence of any sufrace current and surface charges, the normal components of **B** and **D** and the tangential components of **H** and **E** are continuous across an interface.

19.9 Physical Significance of Maxwell's Equations

Let us first consider the equation

---

*More rigorously

$$H_{1t} - H_{2t}$$

is equal to the normal component of the surface current density. However, if there are no surface currents, which is indeed true for most cases, $H_{1t} = H_{2t}$.

$$\operatorname{div} \mathbf{D} = \rho \tag{19.9-1}$$

In free-space

$$\mathbf{D} = \epsilon_0 \mathbf{E} \tag{19.9-2}$$

and Eq. (19.9-1) becomes

$$\operatorname{div} \mathbf{E} = \rho/\epsilon_0 \tag{19.9-3}$$

If we integrate the above equation over a volume V, we obtain,

$$\int \operatorname{div} E \, dV = \frac{1}{\epsilon_0} \int \rho \, dV$$

Applying the divergence theorem we get

$$\oint \mathbf{E} \cdot d\mathbf{a} = \frac{1}{\epsilon_0} Q \tag{19.9-4}$$

which is simply the Gauss' law,* i.e. the electric flux through a closed surface is the total charge inside the volume divided by ϵ_0.

In a similar manner the equation

$$\operatorname{div} \mathbf{B} = 0 \tag{19.9-5}$$

gives

$$\oint \mathbf{B} \cdot d\mathbf{a} = 0 \tag{19.9-6}$$

i.e. the magnetic flux through a closed surface is always zero; this implies the absence of magnetic monopoles.

We next consider the equation

$$\operatorname{curl} \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t} \tag{19.9-7}$$

which associates a space-and time-dependent electric field with a changing magnetic field. Now, Stokes' theorem tells us that

$$\oint_{\Gamma} \mathbf{E} \cdot d\mathbf{l} = \int_{S} \operatorname{curl} \mathbf{E} . d\mathbf{a} \tag{19.9-8}$$

---

*For a dielectric we would get

$$\oint \mathbf{D} . d\mathbf{a} = Q$$

where

$$\mathbf{D} = \epsilon_0 \mathbf{E} + \mathbf{P}$$

P being the dipole moment per unit volume. For a linear homogeneous medium

$$\mathbf{P} = \chi \mathbf{E}$$

where χ is known as the susceptibility. Thus

$$\mathbf{D} = \epsilon \mathbf{E}$$

where

$$\epsilon \equiv \epsilon_0 + \chi$$

is known as the dielectric permittivity of the medium.

where the LHS represents a line integral over a closed path Γ and the RHS represents a surface integral over any surface bounding the path Γ. Thus

$$\oint_{\Gamma} \mathbf{E}\cdot d\mathbf{l} = \int_{S} \text{curl } \mathbf{E}\cdot d\mathbf{a} = -\int_{S} \frac{\partial \mathbf{B}}{\partial t}\, d\mathbf{a} \tag{19.9-9}$$

or,

$$\oint_{\Gamma} \mathbf{E}\cdot dl = -\frac{d}{dt}\int_{S} \mathbf{B}\cdot d\mathbf{a} \tag{19.9-10}$$

where in the last step we have used the fact that the surface S is fixed.* The LHS of the above equation represents the induced emf in a closed circuit which is equal to the negative of the rate of change of the magnetic flux through the circuit. This is the famous Faraday's law of Induction. It is worthwhile to mention that although this law was discovered by Faraday, it was put in the differential form [see Eq. (19.9-8)] by Maxwell.

We now come to the last of the Maxwell's equations i.e.**

$$\text{curl } \mathbf{H} = \mathbf{J} + \frac{\partial \mathbf{D}}{\partial t} \tag{19.9-11}$$

However, Ampere's law (which was known before Maxwell), when expressed as a differential equation was of the form***

$$\text{curl } \mathbf{H} = \mathbf{J} \tag{19.9-12}$$

which implies that a magnetic field is produced only by currents (for example, if we have a long wire carrying a current, we know that it produces a magnetic field). Since the divergence of the curl of any vector is zero, one obtains

$$\text{div } \mathbf{J} = 0 \tag{19.9-13}$$

which may be compared with the equation of continuity

$$\text{div } \mathbf{J} + \frac{\partial \rho}{\partial t} = 0 \tag{19.9-14}$$

Thus Eq. (19.9-13) is valid only when $\frac{\partial \rho}{\partial t} = 0$. Thus, for the Ampere's law to be consistent with the equation of continuity, Maxwell argued that there must be an additional term on the RHS of Eq. (19.9-12)*

---

*Equation (19.9-10) is not valid for a moving system (see, for example, Corson and Lorrain, 1962, p. 526).

**$\mathbf{H} \equiv \frac{\mathbf{B}}{\mu_0} - \mathbf{M}$, where $\mathbf{M}$ is the magnetic moment/unit volume. For a linear material $M = \chi_m \mathbf{H}$ and therefore $\mathbf{B} = \mu \mathbf{H}$, where $\mu = \mu_0 (1 + \chi_m)$.

***Once again, it was Maxwell, who had expressed Ampere's law as a differential equation.

$$\text{curl } \mathbf{H} = \mathbf{J} + \frac{\partial \mathbf{D}}{\partial t}$$

The introduction of the term $\frac{\partial \mathbf{D}}{\partial t}$ (which is known as the displacement current) revolutionized physics. Physically it implies that not only a current produces a magnetic field but a changing electric field also produces a magnetic field (as it indeed happens during the charging and discharging of a condenser).** It may be mentioned that it is the presence of the term $\frac{\partial \mathbf{D}}{\partial t}$ which leads to the wave equation (see Sec. 19.3) and, therefore, the prediction of electromagnetic waves. One can thus argue on physical grounds that a changing electric field produces a magnetic field which varies in space and time and this changing magnetic field produces an electric field varying in space and time, and so on. This mutual generation of electric and magnetic fields result in the propagation of eleotromagnetic waves.

PROBLEMS

1. On the surface of the earth we receive about 1.33 kW of energy per square meter from the sun. Calculate the electric field associated with the sunlight (on the surface of the earth) assuming that it is essentially monochromatic with $\lambda = 6000$ Å. (Ans. ~ 1000 V/m).
2. A 100 W sodium lamp ($\lambda \approx 5890$ Å) is assumed to emit waves uniformly in all directions. What is the radiation pressure on a plane mirror at a distance of 10 m from the bulb.
3. A 1 kW transmitter is emitting electromagnetic waves (of wavelength 40 m) uniformly in all directions. Calculate the electric field at a distance of 1 km from the transmitter.
4. Ocean water can be assumed to be a non-magnetic dielectric with $\kappa \left(= \frac{\epsilon}{\epsilon_0} \right) = 80$ and $\sigma = 4.3$ mhos/m. (a) Calculate the frequency at which the penetration depth will be 10 cm. (b) Show that for frequencies less than 10^8 sec^{-1}, it can be considered as a good conductor. (Ans. (a) $\sim 6 \times 10^5$ sec^{-1}).
5. For silver one may assume $\mu \approx \mu_0$ and $\sigma \approx 3 \times 10^7$ mhos/m. Calculate the skin depth at 10^8 sec^{-1}. (Ans. $\approx 9 \times 10^{-4}$ cm).

---

*Consequently

$$\text{div curl } \mathbf{H} = 0 = \text{div } \mathbf{J} + \frac{\partial}{\partial t} \text{div } \mathbf{D}$$

or

$$0 = \text{div } \mathbf{J} + \frac{\partial \rho}{\partial t}$$

which is the equation of continuity (we have used Eq. 19.9-1).

**For static fields, $\frac{\partial \mathbf{D}}{\partial t} = 0$ and one obtains Ampere's law.

6. Show that for frequencies $\lesssim 10^9$ sec^{-1}, a sample of silicon will act like a good conductor. For silicon one may assume $\frac{\epsilon}{\epsilon_0} \approx 12$ and $\sigma \approx 2$ mhos/cm. Also calculate the penetration depth for this sample at $\nu = 10^6$ sec^{-1}.
7. In a conducting medium show that **H** also satisfies an equation similar to Eq. (19.7–6).
8. Consider an inhomogeneous medium characterised by a space dependent dielectric constant $\epsilon\ (x, y, z)$. Show that

$$\operatorname{div} \mathbf{E} = -\frac{1}{\epsilon}\, \mathbf{E} . \nabla\epsilon$$

Using the results of the previous problem show that the wave equations satisfied by **E** and **H** are

$$\nabla^2 \mathbf{E} + \nabla \left[\mathbf{E} . \frac{1}{\epsilon}\ \nabla\epsilon \right] + \frac{\epsilon\omega^2}{c^2}\, \mathbf{E} = 0$$

$$\nabla^2 \mathbf{H} + \frac{1}{\epsilon}\ (\nabla\epsilon) \times (\nabla \times \mathbf{H}) + \frac{\epsilon\omega^2}{c^2}\, \mathbf{H} = 0$$

Compare the above equations with Eqs. (19.3-8) and (19.3-11) and show that the additional terms appearing in the above equations will be negligible if the variation of ϵ is small in distances of the order of wavelength.

10. Consider the propagation of electromagnetic waves between two parallel perfectly conducting planes. The two surfaces may be assumed to be $y = 0$ and $y = a$. Assume that the fields are independent of x; and the z-and time-dependence to be of the form

$$\exp\left[i\,(\beta z - \omega t)\right]$$

Assuming $E_z = 0$, show that the condition of **E** vanishing on the conducting planes lead to the condition

$$a\sqrt{k^2 - \beta^2} = n\pi; \quad n = 1, 2, 3, \ldots$$

The above equation gives us the allowed values of β. Such a system is known as a waveguide for electromagnetic waves. What are the corresponding field patterns?

11. The above problem corresponds to $E_z = 0$ and the wave is known as TE (Transverse Electric). Solve the problem corresponding to TM (Transverse Magnetic) waves.
12. Consider a layer of dielectric (of refractive index n_1) sandwiched between two dielectrics of higher refractive index n_2. Thus, the refractive index variation is of the form

$$\begin{aligned} n &= n_1 \quad \text{for} \quad |x| < a \\ &= n_2 \quad \text{for} \quad |x| > a \end{aligned}$$

There is no y- (or z-) dependence of n. (Such a slab acts as an optical waveguide and is of extensive use in integrated optics). In order to consider guidance of the wave, assume the z- and time-dependence of **E** and **H** to be of the form:

$$\exp\left[i\,(\beta z - \omega t)\right]$$

Assume TE modes (i.e. $E_z = 0$) with no y-dependence of any field. Show that

$$H_x = -\frac{\beta}{\omega\mu_0}\, E_y \quad \text{and} \quad H_z = -\frac{i}{\omega\mu_0}\, \frac{\partial E_y}{\partial x}$$

where E_y satisfies the following equation

$$\frac{d^2 E_y}{dx^2} + \left[\frac{\omega^2}{c^2}\, n_i^2 - \beta^2\right] E_y = 0; \; i = 1, 2$$

Solve the above equation, use the fact that tangential components of **E** and **H** are continuous and that the fields must tend to zero as $x \to \pm\infty$ to show that

$$\tan pa = \pm \frac{p}{\gamma}$$

where

$$p=\left[\left(\frac{\omega}{c}\,n_1\right)^2-\beta^2\right]^{1/2} \quad \text{and} \quad \gamma=\left[\beta^2-\left(\frac{\omega}{c}\,n_2\right)^2\right]^{1/2}$$

The upper and lower signs correspond to symmetric $[E_y(x) = E_y(-x)]$ and antisymmetric $[E_y(x) = -E_y(-x)]$ modes. The roots of the above equation would give the allowed values of β. (Ref. Sodha and Ghatak, 1977.)

13. Physically interpret the field patterns corresponding to the above problem and show that the waves undergo total internal reflection at the interface and are guided through the system.

REFERENCES AND SUGGESTED READING

See at the end of Chapter 20.

20 Reflection and Refraction of Electromagnetic Waves

But he beholds the light,
and whence it flows
He sees it in his joy

WORDSWORTH

20.1 Introduction

In the previous chapter we had discussed the Maxwell's equations and had shown the existence of electromagnetic waves. We had also shown that at an interface, the tangential components of **E** and **H** and the normal components of **D** and **B** must be continuous. Using these continuity conditions we will, in this chapter, study the reflection and refraction of plane waves at an interface of two dielectrics (Sec. 20.2) and at an interface of a dielectric and a metal (Sec. 20.3).

20.2 Reflection at an Interface of Two Dielectrics

Let us consider the incidence of a plane polarized electromagnetic wave on an interface of two media; we assume the plane $x = 0$ to represent the interface. Let (ϵ_1, μ_1) and (ϵ_2, μ_2) represent the dielectric permittivity and magnetic permeability of the media below and above the plane $x = 0$; we will assume both media to be lossless dielectrics, the case of reflection by a conducting surface will be discussed in Sec. 20.3. Let $\mathbf{E}_1$, $\mathbf{E}_2$ and $\mathbf{E}_3$ denote the electric fields associated with the incident wave, refracted wave and reflected wave respectively. For an incident plane wave, these fields will be of the form

$$\left.\begin{aligned}\mathbf{E}_1 &= \mathbf{E}_{10} \exp\,[i(\mathbf{k}_1\cdot\mathbf{r} - \omega t)]\\ \mathbf{E}_2 &= \mathbf{E}_{20} \exp\,[i(\mathbf{k}_2\cdot\mathbf{r} - \omega_2 t)]\\ \mathbf{E}_3 &= \mathbf{E}_{30} \exp\,[i(\mathbf{k}_3\cdot\mathbf{r} - \omega_3 t)]\end{aligned}\right\} \qquad (20.2\text{-}1)$$

where $\mathbf{E}_{10}$, $\mathbf{E}_{20}$ and $\mathbf{E}_{30}$ are independent of space and time but may, in general, be complex. The vectors $\mathbf{k}_1$, $\mathbf{k}_2$ and $\mathbf{k}_3$ represent the propagation vectors associated with the incident, refracted and reflected waves respectively. Since the fields must satisfy Maxwell's equations we must have (see Sec. 19.2):

$$\left.\begin{aligned} \mathbf{k}_1^2 &= \omega^2 \epsilon_1 \mu_1 \\ \mathbf{k}_2^2 &= \omega_2^2 \epsilon_2 \mu_2 \\ \mathbf{k}_3^2 &= \omega_3^2 \epsilon_1 \mu_1 \end{aligned}\right\} \tag{20.2-2}$$

As discussed in Sec. 19.9, the fields have to satisfy certain boundary conditions at the interface (i.e. at $x = 0$); for example, the tangential component of the electric field must be continuous across it. These conditions are to be satisfied at *all* space points on the interface $x = 0$ and at all times. Consequently, the coefficients of y, z and t in the exponents appearing in Eq. (20.2-1) must be equal. Thus

$$\omega = \omega_2 = \omega_3 \tag{20.2-3}$$

showing that all the waves have the same frequency. Hence Eqs. (20.2-2) simplify to

$$k_1^2 = \omega^2 \epsilon_1 \mu_1 = k_3^2 \tag{20.2-4}$$

$$k_2^2 = \omega^2 \epsilon_2 \mu_2 \tag{20.2-5}$$

Further, if k_{1x}, k_{1y} and k_{1z} represent the x, y and z components of k_1, and similarly for k_2 and k_3, we must have

$$k_{1y} = k_{2y} = k_{3y} \tag{20.2-6}$$

and

$$k_{1z} = k_{2z} = k_{3z} \tag{20.2-7}$$

Without any loss of generality we may choose the y-axis such that

$$k_{1y} = 0$$

(i.e. $\mathbf{k}_1$ is assumed to lie in the x-z plane—see Fig. 20.1). Consequently,

$$k_{2y} = k_{3y} = 0 \tag{20.2-8}$$

Eq. (20.2-8) implies that the vectors $\mathbf{k}_1$, $\mathbf{k}_2$ and $\mathbf{k}_3$ will lie in the same plane. Further, from Eq. (20.2-7) we get

$$k_1 \sin\theta_1 = k_2 \sin\theta_2 = k_3 \sin\theta_3 \tag{20.2-9}$$

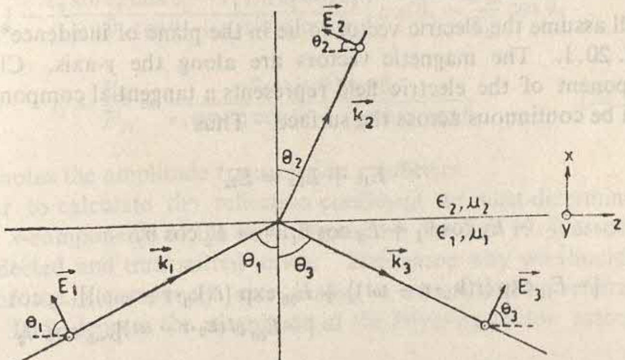


Fig. 20.1 The reflection of a plane wave with its electric vector parallel to the plane of incidence.

Since $k_1 = k_3$ (see Eq. 20.2-4) we must have

$$\theta_1 = \theta_3, \tag{20.2-10}$$

i.e. the angle of incidence is equal to angle of reflection. Further,

$$\frac{\sin \theta_1}{\sin \theta_2} = \frac{k_2}{k_1} = \frac{\omega\sqrt{\epsilon_2\mu_2}}{\omega\sqrt{\epsilon_1\mu_1}},$$

or

$$\frac{\sin \theta_1}{\sin \theta_2} = \sqrt{\frac{\epsilon_2\mu_2}{\epsilon_1\mu_1}} \tag{20.2-11}$$

If $v_1 \left(= \frac{1}{\sqrt{\epsilon_1\mu_1}} \right)$ and $v_2 \left(= \frac{1}{\sqrt{\epsilon_2\mu_2}} \right)$ represent the speeds of propagation of the waves in media 1 and 2, then

$$\frac{\sin \theta_1}{\sin \theta_2} = \frac{v_1}{v_2} = \frac{n_2}{n_1} \tag{20.2-12}$$

where

$$n_1 \left(= \frac{c}{v_1} = c\sqrt{\epsilon_1\mu_1} \right) \text{ and } n_2 \left(= \frac{c}{v_2} = c\sqrt{\epsilon_2\mu_2} \right)$$

represent the refractive indices of media 1 and 2 respectively. Eq. (20.2-12) is the well known Snell's law.

We will now derive expressions for the reflection and transmission coefficients when a plane polarized wave is incident on an interface of two dielectrics. We will first consider the case when the electric vector lies in the plane of incidence which will be followed by the case when the electric vector is at right angles to the plane of incidence.

Case I. E Parallel to the Plane of Incidence

We will assume the electric vector to lie in the plane of incidence* as shown in Fig. 20.1. The magnetic vectors are along the y-axis. Clearly, the z-component of the electric field represents a tangential component which should be continuous across the surface. Thus

$$E_{1z} + E_{3z} = E_{2z}$$

or

$$-E_1 \cos \theta_1 + E_3 \cos \theta_1 = -E_2 \cos \theta_2 \tag{20.2-13}$$

Thus

$$[-E_{10} \exp\{i(\mathbf{k}_1 \cdot \mathbf{r} - \omega t)\} + E_{30} \exp\{i(\mathbf{k}_3 \cdot \mathbf{r} - \omega t)\}]_{x=0} \cos \theta_1 = [-E_{20}\, e^{i(\mathbf{k}_2 \cdot \mathbf{r} - \omega t)}]_{x=0} \cos \theta_2 \tag{20.2-14}$$

Once again, since this condition has to be satisfied at all space points in the plane $x = 0$ and at all times, the exponents must be identically equal which leads to Eq. (20.2-3). Thus

$$[E_{10} - E_{30}] \cos \theta_1 = E_{20} \cos \theta_2 \tag{20.2-15}$$

Further, the normal component of **D** must also be continuous and since $\mathbf{D} = \epsilon \mathbf{E}$, we must have

$$\epsilon_1 E_{1x} + \epsilon_1 E_{3x} = \epsilon_2 E_{2x}$$

or,

$$\epsilon_1 [E_{10} + E_{30}] \sin \theta_1 = \epsilon_2 E_{20} \sin \theta_2 \tag{20.2-16}$$

Substituting for E_{20} from Eq. (20.2-15), we get

$$\epsilon_1 [E_{10} + E_{30}] \sin \theta_1 = \epsilon_2 \sin \theta_2 \frac{[E_{10} - E_{30}]}{\cos \theta_2} \cos \theta_1$$

or,

$$[\epsilon_2 \sin \theta_2 \cos \theta_1 + \epsilon_1 \sin \theta_1 \cos \theta_2] E_{30} = [\epsilon_2 \sin \theta_2 \cos \theta_1 - \epsilon_1 \sin \theta_1 \cos \theta_2] E_{10}$$

Thus

$$r_{\parallel} = \frac{E_{30}}{E_{10}} = \frac{\epsilon_2 \sin \theta_2 \cos \theta_1 - \epsilon_1 \sin \theta_1 \cos \theta_2}{\epsilon_2 \sin \theta_2 \cos \theta_1 + \epsilon_1 \sin \theta_1 \cos \theta_2} \tag{20.2-17}$$

where $r_{\parallel}$ denotes the amplitude reflection coefficient, the subscript $\parallel$ refers to the fact that we are referring to parallel polarization. If we now divide Eq. (20.2-15) by E_{10} and substitute the expression for $\frac{E_{30}}{E_{10}}$ from Eq. (20.2-17) we would get

---

*It is left as an exercise for the readers to show that if the electric vector associated with the incident wave lies in the plane of incidence then the electric vectors associated with the reflected and transmitted waves also lie in the plane of incidence. Similarly, if the electric vector associated with the incident wave is normal to the plane of incidence then the electric vectors associated with the reflected and transmitted waves also lie normal to the plane of incidence.

$$\left[1-\frac{\epsilon_2 \sin\theta_2 \cos\theta_1 - \epsilon_1 \sin\theta_1 \cos\theta_2}{\epsilon_2 \sin\theta_2 \cos\theta_1 + \epsilon_1 \sin\theta_1 \cos\theta_2}\right]\cos\theta_1 = \frac{E_{20}}{E_{10}}\cos\theta_2$$

or,

$$t_{\parallel} = \frac{E_{20}}{E_{10}} = \frac{2\epsilon_1 \sin\theta_1 \cos\theta_1}{\epsilon_2 \sin\theta_2 \cos\theta_1 + \epsilon_1 \sin\theta_1 \cos\theta_2} \qquad (20.2\text{-}18)$$

where $t_{\parallel}$ denotes the amplitude transmission coefficient.

In order to calculate the reflection coefficient we must determine the ratio of the x-components of the Poynting vectors (see Sec. 19.4) associated with the reflected and transmitted waves. The reason why we should take the ratio of the x-component can easily be understood by referring to Fig. 20.2. If S_1 denotes the magnitude of the Poynting vector associated with the incident wave then the energy incident on the area dA (on the surface $x=0$) per unit time would be $S_{1x}\,dA = S_1 dA\cos\theta_1$. Similarly, the energy transmitted through the area dA would be $S_{2x}\,dA = S_2\,dA\cos\theta_2$ and the energy reflected from the area dA would be $S_{3x}\,dA = S_3\cos\theta_1\,dA$. If $R_{\parallel}$ and $T_{\parallel}$ denote the reflection and transmission coefficients then*

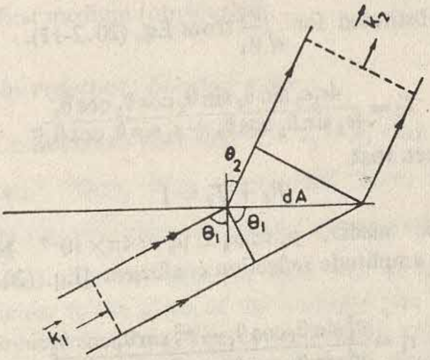


Fig. 20.2 If the cross-sectional area of the incident beam is $dA\cos\theta_1$ then the cross-sectional area of the transmitted beam is $dA\cos\theta_2$.

$$R_{\parallel} = \frac{S_{3x}}{S_{1x}} = \frac{S_3\cos\theta_1}{S_1\cos\theta_1} \qquad (20.2\text{-}19)$$

$$= \frac{<\mathbf{E}_3\times\mathbf{H}_3>}{<\mathbf{E}_1\times\mathbf{H}_1>} = \frac{\sqrt{\frac{\epsilon_1}{\mu_1}}\,|E_{30}|^2}{\sqrt{\frac{\epsilon_1}{\mu_1}}\,|E_{10}|^2} \qquad \text{[See sec. 19.4]}$$

$$= \left|\frac{E_{30}}{E_{10}}\right|^2$$

---

*To calculate the Poynting vector we must use the real parts of E and H; see Sec. 19.4.

or

$$R_\parallel = \left[\frac{\epsilon_2 \sin\theta_2 \cos\theta_1 - \epsilon_1 \sin\theta_1 \cos\theta_2}{\epsilon_2 \sin\theta_2 \cos\theta_1 + \epsilon_1 \sin\theta_1 \cos\theta_2}\right]^2 \qquad (20.2\text{-}20)$$

and

$$T_\parallel = \frac{S_{2x}}{S_{1x}} = \frac{S_2 \cos\theta_2}{S_1 \cos\theta_1}$$

$$= \frac{<\mathbf{E}_2 \times \mathbf{H}_2> \cos\theta_2}{<\mathbf{E}_1 \times \mathbf{H}_1> \cos\theta_1} = \frac{\sqrt{\frac{\epsilon_2}{\mu_2}}\, |E_{20}|^2 \cos\theta_2}{\sqrt{\frac{\epsilon_1}{\mu_1}}\, |E_{10}|^2 \cos\theta_1}$$

$$= \sqrt{\frac{\epsilon_2}{\epsilon_1}} \sqrt{\frac{\epsilon_2}{\epsilon_1}} \frac{\sin\theta_2}{\sin\theta_1} \left[\frac{2\epsilon_1 \sin\theta_1 \cos\theta_1}{\epsilon_2 \sin\theta_2 \cos\theta_1 + \epsilon_1 \sin\theta_1 \cos\theta_2}\right]^2 \frac{\cos\theta_2}{\cos\theta_1}$$

where we have substituted for $\sqrt{\frac{\mu_2}{\mu_1}}$ from Eq. (20.2-11). Thus

$$T_\parallel = \frac{4\epsilon_1\epsilon_2 \sin\theta_1 \sin\theta_2 \cos\theta_1 \cos\theta_2}{[\epsilon_2 \sin\theta_2 \cos\theta_1 + \epsilon_1 \sin\theta_1 \cos\theta_2]^2} \qquad (20.2\text{-}21)$$

It can easily be seen that

$$R_\parallel + T_\parallel = 1 \qquad (20.2\text{-}22)$$

For non-magnetic media, $\mu_1 \approx \mu_2 \approx \mu_0 = 4\pi \times 10^{-7}$ N/amp^2 and the expression for the amplitude reflection coefficient [Eq. (20.2-17)] simplifies to[*]

$$r_\parallel = \frac{n_2^2 \sin\theta_2 \cos\theta_1 - n_1^2 \sin\theta_1 \cos\theta_2}{n_2^2 \sin\theta_2 \cos\theta_1 + n_1^2 \sin\theta_1 \cos\theta_2}$$

But

$$\frac{\sin\theta_1}{\sin\theta_2} = \frac{n_2}{n_1}$$

Therefore,

$$r_\parallel = \frac{\frac{\sin^2\theta_1}{\sin^2\theta_2} \cos\theta_1 \sin\theta_2 - \sin\theta_1 \cos\theta_2}{\frac{\sin^2\theta_1}{\sin^2\theta_2} \cos\theta_1 \sin\theta_2 + \sin\theta_1 \cos\theta_2}$$

$$= \frac{\sin\theta_1 \cos\theta_1 - \sin\theta_2 \cos\theta_2}{\sin\theta_1 \cos\theta_1 + \sin\theta_2 \cos\theta_2} \qquad (20.2\text{-}23)$$

or

$$r_\parallel = \frac{\sin 2\theta_1 - \sin 2\theta_2}{\sin 2\theta_1 + \sin 2\theta_2} = \frac{2\cos(\theta_1 + \theta_2)\sin(\theta_1 - \theta_2)}{2\sin(\theta_1 + \theta_2)\cos(\theta_1 - \theta_2)}$$

$$= \frac{\tan(\theta_1 - \theta_2)}{\tan(\theta_1 + \theta_2)} \qquad (24.2\text{-}24)$$

[*] We are using here the fact that for nonmagnetic media

$$n = c/v = \sqrt{\frac{\epsilon\mu}{\epsilon_0\mu_0}} \approx \sqrt{\frac{\epsilon}{\epsilon_0}}$$

Thus

$$n^2 = \frac{\epsilon}{\epsilon_0}$$

Similarly, starting from Eq. (20.2-18) one easily obtains

$$t_\parallel = \frac{2\cos\theta_1 \sin\theta_2}{\sin(\theta_1+\theta_2)} \tag{20.2-25}$$

From Eqs. (20.2-24) and (20.2-25) we may deduce the following:

(a) *No reflection when $n_2 = n_1$*

When $n_2 = n_1$, $\theta_2 = \theta_1$ and we get

$$r_\parallel = 0 \quad \text{and} \quad t_\parallel = 1$$

Thus there is no reflection when the second medium has the same refractive index as the first medium (obviously!).

(b) *Polarization by reflection*: *Brewster's law*

If the angle of incidence is such that $\theta_1 + \theta_2 = \frac{\pi}{2}$, then $r_\parallel = 0$, i.e. there is no reflected beam.* Thus, if an unpolarized beam is incident at an angle such that $\theta_1 + \theta_2 = \frac{\pi}{2}$, then the parallel component of the E-vector will not be reflected and the reflected light will be polarized with its E-vector perpendicular to the plane of the incidence (see Fig. 20.3). This is the famous *Brewster's Law*. The corresponding angle of incidence is known as the Brewster angle (or the polarizing angle) and is usually

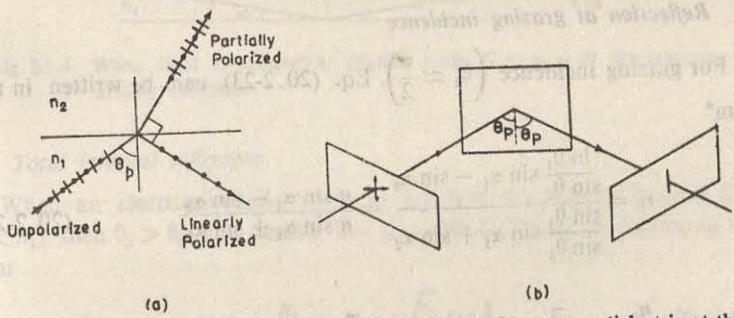


Fig. 20.3 When an unpolarized beam of light is incident on a dielectric at the polarizing angle $\left(\text{i.e. the angle of incidence is equal to } \tan^{-1}\frac{n_2}{n_1}\right)$ then the reflected beam is plane-polarized with its *E*-vector perpendicular to the plane of incidence. The transmitted beam is partially polarized. The dashed line in (b) is the normal to the reflecting surface.

---

*Consequently the entire energy will appear in the transmitted beam. But $t_\parallel = 2\cos^2\theta_1$—why? (See Fig. 20.2).

denoted by θ_p. Notice that the angle of refraction will be $\frac{\pi}{2} - \theta_p$ and therefore Snell's law takes the form

$$\frac{n_2}{n_1} = \frac{\sin\theta_1}{\sin\theta_2} = \frac{\sin\theta_p}{\sin\left(\frac{\pi}{2} - \theta_p\right)} = \tan\theta_p \tag{20.2-26}$$

or

$$\theta_p = \tan^{-1}\left(\frac{n_2}{n_1}\right) \tag{20.2-27}$$

Thus, when the angle of incidence is equal to $\tan^{-1}\left(\frac{n_2}{n_1}\right)$ then the reflected beam is plane polarized. Further, the transmitted beam is partially polarized. It is easily seen that at the polarizing angle, the reflected ray is at right angles to the refracted ray.

(c) *Phase change on reflection and Stokes' relations*

When light is incident on a denser medium, $\theta_2 < \theta_1$ and for $(\theta_1 + \theta_2) > \frac{\pi}{2}$ (i.e. $\theta_1 > \theta_p$), $r_{\parallel}$ is negative implying a phase change of π. However, no such phase change occurs when $\theta_1 < \theta_p$. We will discuss this point in detail later.

The amplitude reflection and transmission coefficients satisfy Stokes' relations (see Example 20.1).

(d) *Reflection at grazing incidence*

For grazing incidence $\left(\theta_1 \approx \frac{\pi}{2}\right)$, Eq. (20.2-23) can be written in the form*

$$r_{\parallel} = \frac{\frac{\sin\theta_1}{\sin\theta_2}\sin\alpha_1 - \sin\alpha_2}{\frac{\sin\theta_1}{\sin\theta_2}\sin\alpha_1 + \sin\alpha_2} = \frac{n\sin\alpha_1 - \sin\alpha_2}{n\sin\alpha_1 + \sin\alpha_2} \tag{20.2-28}$$

where $n = \frac{n_2}{n_1}$, $\alpha_1 = \frac{\pi}{2} - \theta_1$ and $\alpha_2 = \frac{\pi}{2} - \theta_2$ and at grazing incidence both these angles will be small. Now

$$n = \frac{\sin\theta_1}{\sin\theta_2} = \frac{\cos\alpha_1}{\cos\alpha_2}$$

---

*The second medium must be a denser medium (i.e. $n_2 > n_1$) otherwise the beam will undergo total internal reflection [see part (e)].

or

$$\sin \alpha_2 = [1 - \cos^2 \alpha_2]^{1/2} = \left[1 - \frac{\cos^2 \alpha_1}{n^2}\right]^{1/2}$$

Thus

$$r_{\parallel} = \frac{n \sin \alpha_1 - \left[1 - \frac{\cos^2 \alpha_1}{n^2}\right]^{1/2}}{n \sin \alpha_1 + \left[1 - \frac{\cos^2 \alpha_1}{n^2}\right]^{1/2}} \approx \frac{n\alpha_1 - \left[1 - \frac{1}{n^2}\right]^{1/2}}{n\alpha_1 + \left[1 - \frac{1}{n^2}\right]^{1/2}} \tag{20.2-29}$$

where we have replaced sin α_1 by α_1 and cos α_1 by 1 (thus we have retained terms proportional to α_1 but neglected terms of higher order—this will be justified when α_1 is small). Thus

$$r_{\parallel} \approx -\left[1 - \frac{n\alpha_1}{\sqrt{\frac{n^2-1}{n^2}}}\right]\left[1 + \frac{n\alpha_1}{\sqrt{\frac{n^2-1}{n^2}}}\right]^{-1}$$

$$\approx -\left[1 - \frac{2n^2 \alpha_1}{\sqrt{n^2-1}}\right] \rightarrow -1 \text{ as } \alpha_1 \rightarrow 0 \tag{20.2-30}$$

which shows that the reflection is complete at grazing incidence. The transmission coefficient tends to zero as it is indeed obvious from Eq. (20.2-25). Thus, if we hold a glass plate horizontally at the level of the eye (see Fig. 20.4) the angle of incidence will be close to $\pi/2$ and the plate will act as a mirror.

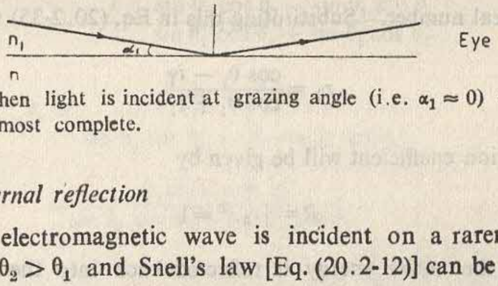


Fig. 20.4 When light is incident at grazing angle (i.e. $\alpha_1 \approx 0$) the reflection is almost complete.

(e) *Total internal reflection*

When an electromagnetic wave is incident on a rarer medium (i.e. $n_2 < n_1$) then $\theta_2 > \theta_1$ and Snell's law [Eq. (20.2-12)] can be written in the form

$$\sin \theta_2 = \frac{n_1}{n_2} \sin \theta_1 = \sqrt{\frac{\epsilon_1}{\epsilon_2}} \sin \theta_1 \tag{20.2-31}$$

where the media have been assumed to be non-magnetic, i.e. we have assumed

$$n_1 = \sqrt{\frac{\epsilon_1}{\epsilon_2}} \text{ and } n_2 = \sqrt{\frac{\epsilon_2}{\epsilon_0}} \tag{20.2-32}$$

Clearly

$$\sin \theta_2 > 1$$

when

$$\theta_1 > \theta_c$$

where

$$\theta_c = \sin^{-1}\left(\frac{n_2}{n_1}\right) = \sin^{-1}\sqrt{\frac{\epsilon_2}{\epsilon_1}} \tag{20.2-33}$$

The angle θ_c is known as the critical angle. Now, for non-magnetic media the amplitude reflection coefficient will be given by [see Eq. (20.2-17)]:

$$\begin{aligned} r_{\parallel} &= \frac{\epsilon_2 \cos\theta_1 \sin\theta_2 - \epsilon_1 \sin\theta_1 \cos\theta_2}{\epsilon_2 \cos\theta_1 \sin\theta_2 + \epsilon_1 \sin\theta_1 \cos\theta_2} \\ &= \frac{\cos\theta_1 - \sqrt{\epsilon_1/\epsilon_2}\,\sqrt{1-\sin^2\theta_2}}{\cos\theta_1 + \sqrt{\epsilon_1/\epsilon_2}\,\sqrt{1-\sin^2\theta_2}} \end{aligned} \tag{20.2-34}$$

$$= \frac{\cos\theta_1 - \frac{\epsilon_1}{\epsilon_2}\sqrt{\frac{\epsilon_2}{\epsilon_1} - \sin^2\theta_1}}{\cos\theta_1 + \frac{\epsilon_1}{\epsilon_2}\sqrt{\frac{\epsilon_2}{\epsilon_1} - \sin^2\theta_1}} = \frac{\cos\theta_1 - \frac{\epsilon_1}{\epsilon_2}\sqrt{\sin^2\theta_c - \sin^2\theta_1}}{\cos\theta_1 + \frac{\epsilon_1}{\epsilon_2}\sqrt{\sin^2\theta_c - \sin^2\theta_1}} \tag{20.2-35}$$

where we have used Eqs. (20.2-31) and (20.2-33). Clearly, for $\theta_1 > \theta_c$ the quantity under the square root becomes negative and we may write

$$\frac{\epsilon_1}{\epsilon_2}\sqrt{\sin^2\theta_c - \sin^2\theta_1}\left(= \frac{\epsilon_1}{\epsilon_2}\sqrt{\frac{\epsilon_2}{\epsilon_1} - \sin^2\theta_1}\right) = i\gamma \tag{20.2-36}$$

where γ is a real number. Substituting this in Eq. (20.2-35) we get

$$r_{\parallel} = \frac{\cos\theta_1 - i\gamma}{\cos\theta_1 + i\gamma}, \tag{20.2-37}$$

and the reflection coefficient will be given by

$$R = |r_{\parallel}|^2 = 1 \tag{20.2-38}$$

showing that the entire energy is reflected back into the first medium. This is the well known phenomenon of 'total internal reflection'. We may, however, note two points:

(i) since $r_{\parallel}$ is a complex number, the phase change on reflection lies between 0 and π (see Example 20.3).

(ii) The amplitude transmission coefficient is given by

$$t_{\parallel} = \frac{2\epsilon_1 \sin\theta_1 \cos\theta_1}{\epsilon_1 \sin\theta_1 \cos\theta_2 + \epsilon_2 \cos\theta_1 \sin\theta_2}$$

which is not zero. Thus the field in the rarer medium is not zero (see Example 20.4).

Example 20.1 Fig 20.5 shows that if the media are interchanged the angles of incidence and refraction are reversed. If $r_{\parallel}'$ and $t_{\parallel}'$ denote the

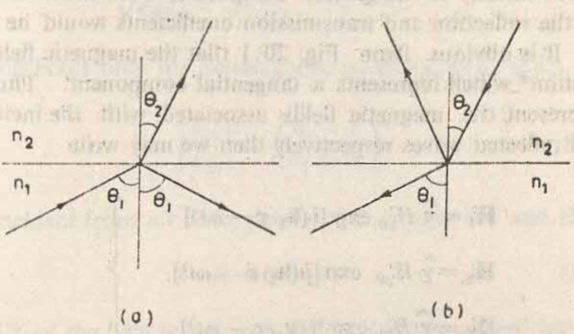


Fig. 20.5 The angles of incidence and refraction are reversed if the media are interchanged.

amplitude reflection and transmission coefficients corresponding to Fig. 20.5(b), then show that

$$1 + r_{\parallel} r_{\parallel}' = t_{\parallel} t_{\parallel}' \tag{20.2-39}$$

(This is one of the Stokes relation—see Chapter 11).

Solution.

$$r_{\parallel} = \frac{\sin\theta_1 \cos\theta_1 - \sin\theta_2 \cos\theta_2}{\sin\theta_1 \cos\theta_1 + \sin\theta_2 \cos\theta_2} \tag{20.2-40}$$

(see Eq. 20.2-23)

$$r_{\parallel}' = \frac{\sin\theta_2 \cos\theta_2 - \sin\theta_1 \cos\theta_1}{\sin\theta_2 \cos\theta_2 + \sin\theta_1 \cos\theta_1} = -r_{\parallel} \tag{20.2-41}$$

(see Fig. 20.5)

$$t_{\parallel} = \frac{2\sin\theta_1 \cos\theta_1}{\sin\theta_1 \cos\theta_2 + \frac{\epsilon_2}{\epsilon_1} \cos\theta_1 \sin\theta_2}$$

$$= \frac{2\sin\theta_1 \cos\theta_1}{\sin\theta_2 \cos\theta_2 + \sin\theta_1 \cos\theta_1} \qquad \left(\text{because } \frac{\epsilon_2}{\epsilon_1} = \frac{n_2^2}{n_1^2} = \frac{\sin^2\theta_1}{\sin^2\theta_2}\right)$$

and

$$t_{\parallel}' = \frac{2\cos\theta_2 \sin\theta_2}{\sin\theta_1 \cos\theta_1 + \sin\theta_2 \cos\theta_2}$$

Simple substitutions give

$$1 + r_{\parallel} r_{\parallel}' = 1 - |r_{\parallel}|^2 = 1 - \left[\frac{\sin\theta_1 \cos\theta_1 - \sin\theta_2 \cos\theta_2}{\sin\theta_1 \cos\theta_1 + \sin\theta_2 \cos\theta_2}\right]^2$$

$$= t_{\parallel}^2$$

Example 20.2 In deriving the reflection and transmission coefficients.

instead of assuming the continuity of the normal component of **D**, if we assume the continuity of tangential component of **H** show that the same results for the reflection and transmission coefficients would be obtained.

Solutions. It is obvious from Fig. 20.1 that the magnetic field will be in the y-direction* which represents a tangential component. Thus if $\mathbf{H}_1$, $\mathbf{H}_2$ and $\mathbf{H}_3$ represent the magnetic fields associated with the incident, transmitted and reflected waves respectively then we may write

$$\left.\begin{aligned} \mathbf{H}_1 &= \hat{\mathbf{y}}\, H_{10} \exp\left[i\left(\mathbf{k}_1\cdot\mathbf{r} - \omega t\right)\right] \\ \mathbf{H}_2 &= \hat{\mathbf{y}}\, H_{20} \exp\left[i\left(\mathbf{k}_2\cdot\mathbf{r} - \omega t\right)\right] \\ \mathbf{H}_3 &= \hat{\mathbf{y}}\, H_{30} \exp\left[i\left(\mathbf{k}_3\cdot\mathbf{r} - \omega t\right)\right] \end{aligned}\right\} \qquad (20.2\text{-}42)$$

Continuity of the y-component of the field would give

$$H_{10} + H_{30} = H_{20} \qquad (20.2\text{-}43)$$

But

$$\mathbf{H} = \frac{\mathbf{k}\times\mathbf{E}}{\omega\mu} \qquad (20.2\text{-}44)$$

Thus

$$\frac{k_1}{\omega\mu_1}\left(E_{10} + E_{30}\right) = \frac{k_2}{\omega\mu_2} E_{20} \qquad (20.2\text{-}45)$$

Continuity of the tangential component of **E** gave (see Eq. 20.2-15)

$$\begin{aligned} \left(E_{10} - E_{30}\right)\cos\theta_1 &= E_{20}\cos\theta_2 \\ &= \frac{k_1}{k_2}\frac{\mu_2}{\mu_1}\left(E_{10} + E_{30}\right)\cos\theta_2 \end{aligned}$$

or

$$r_{\parallel} = \frac{E_{30}}{E_{10}} = \frac{\dfrac{k_2}{\mu_2}\cos\theta_1 - \dfrac{k_1}{\mu_1}\cos\theta_2}{\dfrac{k_2}{\mu_2}\cos\theta_1 + \dfrac{k_1}{\mu_1}\cos\theta_2} \qquad (20.2\text{-}46)$$

$$= \frac{\sqrt{\dfrac{\epsilon_2}{\mu_2}}\cos\theta_1 - \sqrt{\dfrac{\epsilon_1}{\mu_1}}\cos\theta_2}{\sqrt{\dfrac{\epsilon_2}{\mu_2}}\cos\theta_1 + \sqrt{\dfrac{\epsilon_1}{\mu_1}}\cos\theta_2} \qquad (20.2.47)$$

If we now use Snell's law, i.e.

$$\frac{\sin\theta_1}{\sin\theta_2} = \sqrt{\frac{\epsilon_2\mu_2}{\epsilon_1\mu_1}}$$

we would get Eq. (20.2-17). However, from Eq. (20.2-35) we get the reflection coefficient at normal incidence

---

*The vector $\mathbf{E}\times\mathbf{H}$ is along the direction of propagation.

$$r_{\parallel}=\frac{\sqrt{\epsilon_2/\mu_2}-\sqrt{\epsilon_1/\mu_1}}{\sqrt{\epsilon_2/\mu_2}+\sqrt{\epsilon_1/\mu_1}}\approx\frac{n_2-n_1}{n_2+n_1} \tag{20.2-48}$$

the last relation holds only for nonmagnetic media ($\mu_2 \approx \mu_1 \approx \mu_0$). Thus

$$R=\left|\frac{E_{30}}{E_{10}}\right|^2=\left(\frac{n_2-n_1}{n_2+n_1}\right)^2 \tag{20.2-49}$$

For a beam incident from air onto glass $n_1 = 1.0$, $n_2 = 1.5$ and therefore,

$$R = 0.04 \tag{20.2-50}$$

Thus about 4% of the light is reflected and 96% is transmitted into glass.

Example 20.3 Calculate the phase change in the beam which undergoes total internal reflection.

Solution.

$$\begin{aligned} r_{\parallel} &= \frac{\cos\theta_1 - i\gamma}{\cos\theta_1 + i\gamma} \\ &= \frac{Ae^{-i\phi}}{Ae^{i\phi}} = e^{-2i\phi} \end{aligned} \tag{20.2-51}$$

where

$$A = [\cos^2\theta_1 + \gamma^2]^{1/2}$$

$$\cos\phi = \frac{\cos\theta_1}{[\cos^2\theta + \gamma^2]^{1/2}}, \quad \sin\phi = \frac{\gamma}{[\cos^2\theta_1 + \gamma^2]^{1/2}}$$

Thus

$$E_{30} = E_{10}\, e^{-2i\phi} \tag{20.2-52}$$

and the phase change (Δ) is given by

$$\Delta = 2\phi = 2\tan^{-1}\frac{\gamma}{\cos\theta_1} = 2\tan^{-1}\frac{\epsilon_1}{\epsilon_2}\frac{\sqrt{\sin^2\theta_1 - \sin^2\theta_0}}{\cos\theta_1} \tag{20.2-53}$$

Example 20.4 Determine the nature of the transmitted wave when the beam undergoes total internal reflection.

Solution. The electric field associated with the transmitted wave is given by (see Eq. 20.2-1):

$$\begin{aligned} E_2 &= \mathbf{E}_{20}\exp\,[i\,(\mathbf{k}_2\cdot\mathbf{r} - \omega t)] \\ &= \mathbf{E}_{20}\exp\,[i\,(\mathbf{k}_{2x}\,x + k_{2z}\,z - \omega t)] \\ &= \mathbf{E}_{20}\exp\,[i\,(k_2\,x\cos\theta_2 + k_2\,z\sin\theta_2 - \omega t)] \end{aligned} \tag{20.2-54}$$

(see Fig. 20.1)

Now

$$\frac{\sin\theta_1}{\sin\theta_2} = \sqrt{\frac{\epsilon_2}{\epsilon_1}}$$

$$\therefore \quad \sin\theta_2 = \sqrt{\epsilon_1/\epsilon_2}\,\sin\theta_1$$

and

$$\cos\theta_2 = \sqrt{1 - \frac{\epsilon_1}{\epsilon_2}\sin^2\theta_1} = \sqrt{\frac{\epsilon_1}{\epsilon_2}}\sqrt{\frac{\epsilon_2}{\epsilon_1} - \sin^2\theta_1}$$

$$= \sqrt{\frac{\epsilon_1}{\epsilon_2}}\, i\gamma$$

Thus

$$\mathbf{E}_2 = \mathbf{E}_{20}\exp(-\beta x)\exp\left[i\left\{\left(k_2\sqrt{\frac{\epsilon_1}{\epsilon_2}}\sin\theta_1\right)z - \omega t\right\}\right] \quad (20.2\text{-}55)$$

where

$$\beta = k_2\sqrt{\frac{\epsilon_2}{\epsilon_1}}\,\gamma \quad (20.2\text{-}56)$$

The field given by Eq. (20.2-55) represents a wave propagating in the $+z$-direction with an amplitude decreasing exponentially in the x-direction. Such a wave is known as a 'surface-wave' or an 'evanescent wave' (see Fig. 20.6). Such waves have many interesting applications.*

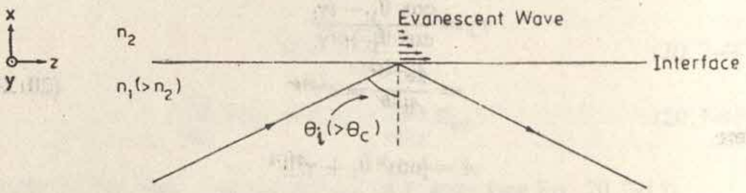


Fig. 20.6 An evanescent wave is generated in the rarer medium when a beam undergoes total internal reflection. The evanescent wave is propagating along the z-axis and the amplitude decreases along the x-axis.

Case 2. E Perpendicular to the Plane of Incidence

Let us next consider the reflection and refraction of a linearly polarized plane wave with its electric vector perpendicular to the plane of incidence; the reflection is assumed to occur at the interface of two dielectrics. Thus the electric vectors will be along the y-axis (see Fig. 20.7) and we may write

$$\left.\begin{aligned} \mathbf{E}_1 &= \hat{\mathbf{y}}\, E_{10}\exp[i(\mathbf{k}_1\cdot\mathbf{r} - \omega t)] \\ \mathbf{E}_2 &= \hat{\mathbf{y}}\, E_{20}\exp[i(\mathbf{k}_2\cdot\mathbf{r} - \omega t)] \\ \mathbf{E}_3 &= \hat{\mathbf{y}}\, E_{30}\exp[i(\mathbf{k}_3\cdot\mathbf{r} - \omega t)] \end{aligned}\right\} \quad (20.2\text{-}57)$$

where $\mathbf{E}_1$, $\mathbf{E}_2$ and $\mathbf{E}_3$ denote the electric vectors associated with the incident, transmitted and reflected waves respectively. Since the y-axis is tangential to the interface, the y-component of E must be continuous across the interface; consequently

---

*See, for example, Bryngdahl (1973).

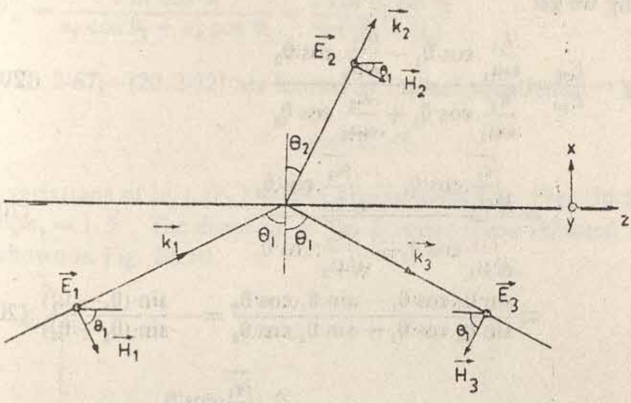


Fig. 20.7 The reflection and refraction of a plane wave, the electric vector lies perpendicular to the plane of incidence.

$$E_{10} + E_{30} = E_{20} \tag{20.2-58}$$

The directions of the magnetic fields* are also shown in Fig. 20.7; they lie in the plane of incidence and are given by

$$\left.\begin{aligned} \mathbf{H}_1 &= \mathbf{H}_{10} \exp[i(\mathbf{k}_1 \cdot \mathbf{r} - \omega t)] = \frac{\mathbf{k}_1 \times \mathbf{E}_{10}}{\omega\mu_1} \exp[i(\mathbf{k}_1 \cdot \mathbf{r} - \omega t)] \\ \mathbf{H}_2 &= \mathbf{H}_{20} \exp[i(\mathbf{k}_2 \cdot \mathbf{r} - \omega t)] = \frac{\mathbf{k}_2 \times \mathbf{E}_{20}}{\omega\mu_2} \exp[i(\mathbf{k}_2 \cdot \mathbf{r} - \omega t)] \\ \mathbf{H}_3 &= \mathbf{H}_{30} \exp[i(\mathbf{k}_3 \cdot \mathbf{r} - \omega t)] = \frac{\mathbf{k}_3 \times \mathbf{E}_{30}}{\omega\mu_1} \exp[i(\mathbf{k}_3 \cdot \mathbf{r} - \omega t)] \end{aligned}\right\} \tag{20.2-59}$$

(Notice that **H** lies in the plane of incidence). Since $\mathbf{k}_1$ is at right angles to $\mathbf{E}_{10}$, the magnitude of $\mathbf{H}_{10}$ is simply $\frac{k_1 E_{10}}{\omega\mu_1}$; similarly for $\mathbf{H}_{20}$ and $\mathbf{H}_{30}$. It is obvious from Fig. 20.7 that for the z-component of the magnetic field to be continuous, we must have

$$H_{10} \cos\theta_1 - H_{30} \cos\theta_1 = H_{20} \cos\theta_2 \tag{20.2-60}$$

or

$$\frac{k_1}{\omega\mu_1}(E_{10} - E_{30}) \cos\theta_1 = \frac{k_2}{\omega\mu_2} E_{20} \cos\theta_2 \tag{20.2-61}$$

Substituting the expression for E_{20} from Eq. (20.2-58) we get

$$\frac{k_1}{\omega\mu_1}(E_{10} - E_{30}) \cos\theta_1 = \frac{k_2}{\omega\mu_2}(E_{10} + E_{30}) \cos\theta_2$$

---

*It may be noted that since the displacement vector **D** has no component normal to the interface, the continuity of normal component of **D** will not give us any equation.

Rearranging we get

$$r_\perp = \frac{E_{30}}{E_{10}} = \frac{\dfrac{k_1}{\omega\mu_1}\cos\theta_1 - \dfrac{k_2}{\omega\mu_2}\cos\theta_2}{\dfrac{k_1}{\omega\mu_1}\cos\theta_1 + \dfrac{k_2}{\omega\mu_2}\cos\theta_2} \qquad (20.2\text{-}62)$$

$$= \frac{\sqrt{\dfrac{\epsilon_1}{\mu_1}}\cos\theta_1 - \sqrt{\dfrac{\epsilon_2}{\mu_2}}\cos\theta_2}{\sqrt{\dfrac{\epsilon_1}{\mu_1}}\cos\theta_1 + \sqrt{\dfrac{\epsilon_2}{\mu_2}}\cos\theta_2} \qquad (20.2\text{-}63)$$

$$\approx \frac{\sin\theta_2\cos\theta_1 - \sin\theta_1\cos\theta_2}{\sin\theta_2\cos\theta_1 + \sin\theta_1\cos\theta_2} = -\frac{\sin(\theta_1-\theta_2)}{\sin(\theta_1+\theta_2)} \qquad (20.2\text{-}64)$$

Further

$$t_\perp = \frac{E_{20}}{E_{10}} = 1 + \frac{E_{30}}{E_{10}} = \frac{2\sqrt{\dfrac{\epsilon_1}{\mu_1}}\cos\theta_1}{\sqrt{\dfrac{\epsilon_1}{\mu_1}}\cos\theta_1 + \sqrt{\dfrac{\epsilon_2}{\mu_2}}\cos\theta_2} \qquad (20.2\text{-}65)$$

$$\approx \frac{2\sin\theta_2\cos\theta_1}{\sin(\theta_1+\theta_2)} \qquad (20.2\text{-}66)$$

where the subscript $\perp$ on r and t refers to the fact that we are referring to the state of polarization in which the E-vector is perpendicular to the plane of incidence. It may be mentioned that Eqs. (20.2-62), (20.2-63) and (20.2-65) are exact, whereas Eqs. (20.2-64) and (20.2-66) are valid for non-magnetic media. Once again, one can show that when $\theta_1 > \theta_c$, total internal reflection will occur and for grazing incidence the reflection is complete.

We may summarize now the amplitude reflection and transmission coefficients for the two cases; the results being valid for non-magnetic media:

$$r_\parallel = \frac{n_2\cos\theta_1 - n_1\cos\theta_2}{n_2\cos\theta_1 + n_1\cos\theta_2} = \frac{\left(\dfrac{n_2}{n_1}\right)^2\cos\theta_1 - \sqrt{\left(\dfrac{n_2}{n_1}\right)^2 - \cos^2\theta_1}}{\left(\dfrac{n_2}{n_1}\right)^2\cos\theta_1 + \sqrt{\left(\dfrac{n_2}{n_1}\right)^2 - \cos^2\theta_1}} \qquad (20.2\text{-}67)$$

$$= \frac{\tan(\theta_1-\theta_2)}{\tan(\theta_1+\theta_2)} \qquad (20.2\text{-}68)$$

$$r_\perp = \frac{n_1\cos\theta_1 - n_2\cos\theta_2}{n_1\cos\theta_1 + n_2\cos\theta_2} = \frac{\cos\theta_1 - \sqrt{\left(\dfrac{n_2}{n_1}\right)^2 - \cos^2\theta_1}}{\cos\theta_1 + \sqrt{\left(\dfrac{n_2}{n_1}\right)^2 - \cos^2\theta_1}} \qquad (20.2\text{-}69)$$

$$= -\frac{\sin(\theta_1-\theta_2)}{\sin(\theta_1+\theta_2)} \qquad (20.2\text{-}70)$$

$$t_\parallel = \frac{2\,n_1\cos\theta_1}{n_2\cos\theta_1 + n_1\cos\theta_2} = \frac{2\cos\theta_1\sin\theta_2}{\sin\theta_1\cos\theta_1 + \sin\theta_2\cos\theta_2} \qquad (20.2\text{-}71)$$

$$t_\perp = \frac{2\, n_1 \cos \theta_1}{n_1 \cos \theta_1 + n_2 \cos \theta_2} = \frac{2 \cos \theta_1 \sin \theta_2}{\sin (\theta_1 + \theta_2)} \tag{20.2-72}$$

Eq. (20.2-67)—(20.2-72) are known as Fresnel equations.* We write

$$r = |r|\, e^{i\phi}$$

The variations of $|r_\parallel|$, $|r_\perp|$, $\phi_\parallel$ and $\phi_\perp$ are plotted in Figs. 20.8 and 20.9 for $n_2/n_1 = 1.5$. The direction of the E-vector in the reflected components are shown in Fig. 20.10.

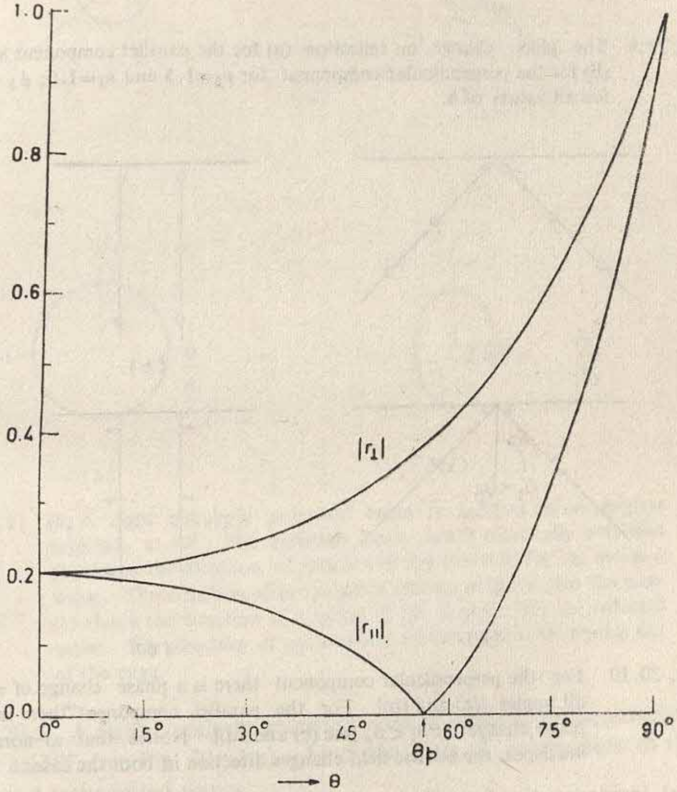


Fig. 20.8 Variation of $|r_\parallel|$ and $|r_\perp|$ with the angle of incidence when $n_2 = 1.5$ and $n_1 = 1.0$

Referring to Fig. 20.8 we note that when $\theta_1 = \tan^{-1}(n_2/n_1) \approx 56°$, $|r_\parallel| = 0$. This is the Brewster's angle. At grazing incidence (i.e. as $\theta_1 \to 90°$), both $|r_\parallel|$ and $|r_\perp|$ tend to 1 implying complete reflection. At

---

*An alternative derivation of Fresnel equations has been given by Feynmann, *et al.* (1964)—Vol. I § 33-6.

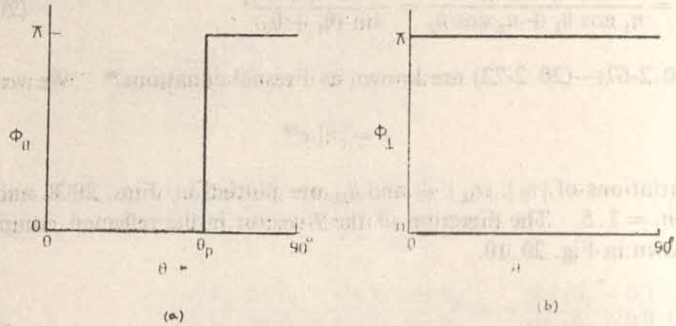


Fig. 20.9 The phase change on reflection (a) for the parallel component and (b) for the perpendicular component for $n_2=1.5$ and $n_1=1.0$; $\phi_{\perp}=\pi$ for all values of θ.

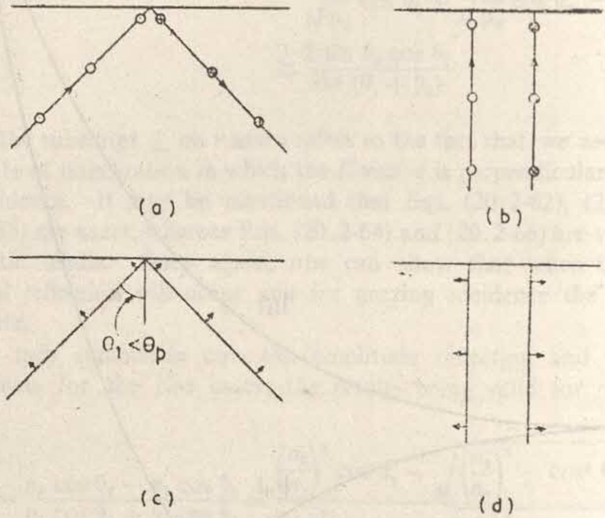


Fig. 20.10 For the perpendicular component there is a phase change of π at all angles [(a) and (b)]. For the parallel component there is no phase change for $\theta_1 < \theta_p$ [see (c) and (d)]. Notice that at normal incidence, the electric field changes direction in both the cases.

normal incidence (i.e. $\theta_1 = 0$) any state of polarization can be thought of as parallel polarization or perpendicular polarization* and we should expect $r_{\parallel}$ and $r_{\perp}$ to give the same result. Fig. 20.8 shows that both $|r_{\parallel}|$ and $|r_{\perp}|$ have the same value; however Fig. 20.9 shows that whereas the perpendicular component predicts a phase change of π, there is no phase change associated with the parallel component. There is, however, no

---

*This is due to the fact that at normal incidence the direction of propagation is coincident with the normal to the reflecting surface and *any* plane containing the normal could be thought of as the plane of incidence.

inconsistency, if we study the direction of the electric vector associated with the reflected component [see (b) and (d) of Fig. 20.10].

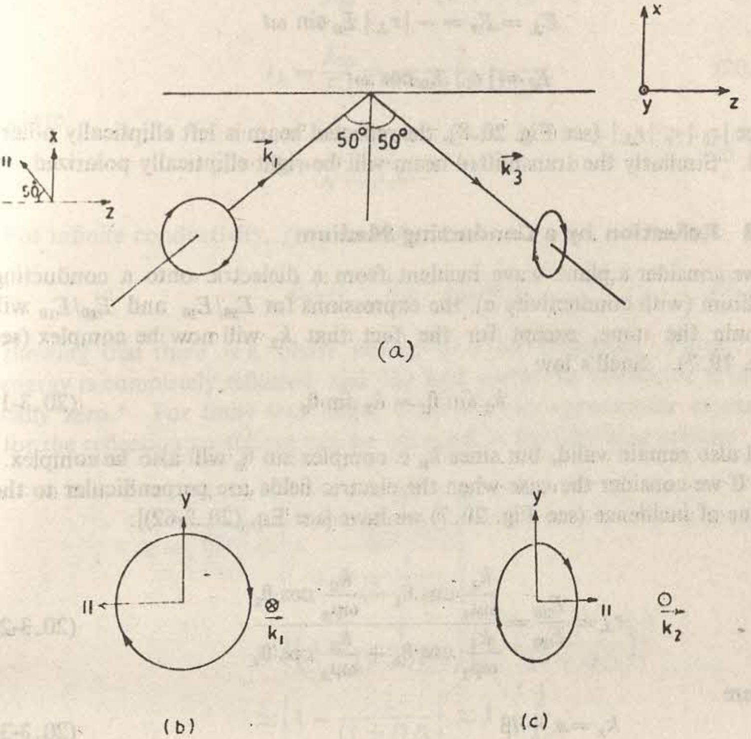


Fig. 20.11 (a) A right circularly polarized beam is incident on an air-glass interface at 50°. The reflected beam is left elliptically polarized (b) shows the direction of rotation of the E-vector for the incident wave. The direction of propagation (shown as ⊗) is into the page (c) shows the direction of rotation of the E-vector for the reflected wave. The direction of propagation (shown as ⊙) is coming out of the page.

Example 20.5 A right circularly polarized wave is incident on an air-glass interface at 50° (see Fig. 20.11 a). Find the state of polarization of the reflected and transmitted waves.

Solution. For a right circularly polarized wave (Fig. 20.11 b) we may write

$$E_{\perp} = E_y = E_0 \sin \omega t$$

$$E_{\parallel} = E_0 \cos \omega t;$$

$$E_z = - E_{\parallel} \cos \theta_1 = - E_0 \sin \omega t \cos 50$$

$$E_x = E_{\parallel} \sin \theta_1 = E_0 \sin \omega t \sin 50$$

Since the angle of incidence is less than the Brewster angle, the electric fields associated with the reflected wave would be

$$E_{\perp} = E_y = -|r_{\perp}|\, E_0 \sin \omega t$$

$$E_{\parallel} = |r_{\parallel}|\, E_0 \cos \omega t$$

Since $|r_{\parallel}| < |r_{\perp}|$ (see Fig. 20.8), the reflected beam is left elliptically polarized. Similarly the transmitted beam will be right elliptically polarized.

20.3 Reflection by a Conducting Medium

If we consider a plane wave incident from a dielectric onto a conducting medium (with conductivity σ), the expressions for E_{20}/E_{10} and E_{30}/E_{10} will remain the same, except for the fact that k_2 will now be complex (see Sec. 19.7). Snell's law

$$k_1 \sin \theta_1 = k_2 \sin \theta_2 \tag{20.3-1}$$

will also remain valid, but since k_2 is complex $\sin \theta_2$ will also be complex.

If we consider the case when the electric fields are perpendicular to the plane of incidence (see Fig. 20.7) we have [see Eq. (20.2-62)]:

$$r_{\perp} = \frac{E_{30}}{E_{10}} = \frac{\dfrac{k_1}{\omega\mu_1}\cos\theta_1 - \dfrac{k_2}{\omega\mu_2}\cos\theta_2}{\dfrac{k_1}{\omega\mu_1}\cos\theta_1 + \dfrac{k_2}{\omega\mu_2}\cos\theta_2} \tag{20.3-2}$$

where

$$k_2 = \alpha + i\beta \tag{20.3-3}$$

$$\alpha = \omega\sqrt{\epsilon_2\mu_2}\left[\tfrac{1}{2} + \tfrac{1}{2}\sqrt{1 + \frac{\sigma^2}{\omega^2\epsilon_2^2}}\right]^{1/2}, \quad \beta = \frac{\omega\sigma\mu_2}{2\alpha} \tag{20.3-4}$$

(see Sec. 19.7). For normal incidence

$$r_{\perp} = \frac{E_{30}}{E_{10}} = \frac{1 - \dfrac{(\alpha + i\beta)}{\omega\sqrt{\epsilon_1\mu_1}}\dfrac{\mu_1}{\mu_2}}{1 + \dfrac{(\alpha + i\beta)}{\omega\sqrt{\epsilon_1\mu_1}}\dfrac{\mu_1}{\mu_2}} \tag{20.3-5}$$

and

$$t_{\perp} = \frac{E_{20}}{E_{10}} = 1 + r_{\perp} = \frac{2}{1 + \dfrac{(\alpha + i\beta)\,\mu_1}{\omega\sqrt{\epsilon_1\mu_1}\,\mu_2}} \tag{20.3-6}$$

For a good conductor $\dfrac{\sigma}{\epsilon\omega} \gg 1$ and

$$\alpha \approx \beta \approx \sqrt{\omega\sigma\mu_2/2}$$

Thus

$$r_\perp = \frac{E_{30}}{E_{10}} = \frac{1-(1+i)\Delta}{1+(1+i)\Delta} \tag{20.3-8}$$

$$t_\perp = \frac{E_{20}}{E_{10}} = \frac{2}{1+(1+i)\Delta} \tag{20.3-9}$$

where

$$\Delta = \sqrt{\frac{\sigma\mu_1}{2\mu_2\epsilon_1\omega}} \tag{20.3-10}$$

For infinite conductivity, $\Delta \to \infty$ and

$$E_{30} = -E_{10}\,, \; E_{20} = 0; \tag{20.3-11}$$

showing that there is a phase change of π on reflection. Further, the energy is completely reflected and the field inside the conductor is identically zero.* For finite (but large) value of σ, an approximate expression for the reflection coefficient can be obtained in the following manner:

$$\begin{aligned} R = \left|\frac{E_{30}}{E_{10}}\right|^2 &= \left| - \frac{1-\dfrac{1}{(1+i)\Delta}}{1+\dfrac{1}{(1+i)\Delta}} \right|^2 \\ &\approx \left| \left(1-\frac{1}{(1+i)\Delta}\right)\left(1-\frac{1}{(1+i)\Delta}\right) \right|^2 \\ &\approx \left| 1-\frac{2}{(1+i)\Delta} \right|^2 \approx 1-\frac{2}{\Delta} \\ &\approx 1-2\sqrt{\frac{2\mu_2\epsilon_1\omega}{\sigma\mu_1}} \end{aligned} \tag{20.3-12}$$

For non-magnetic media, with

$$\epsilon_1 \approx \epsilon_0 = 8.854\times10^{-12}\ \mathrm{C^2/Nm^2}$$

$$\omega \approx 2\pi\times10^{10}\ \mathrm{sec}^{-1},\ \sigma \approx 3\times10^{7}\ \text{mhos/m (silver)},\ R \approx 0.9996.$$

Thus about 99.96% of light is reflected. This is the reason why metals are such good reflectors. Notice that the reflection coefficient increases with decrease in frequency.

When the incidence is not normal one must substitute the following expression for $\cos\theta_2$:

---

*It should be noted that if $\sigma \to \infty$ (i.e. for a perfect conductor, $r_\parallel \to +1$ and $r_\perp \to -1$ (see Eqs. 20.2-46 and 20.3-2) even for non-normal incidence. Thus, if a right circularly polarized wave is incident on a perfect conductor then the reflected light will be left circularly polarized.

$$\begin{aligned}\cos\theta_2 &= \sqrt{1-\sin^2\theta_2} = \sqrt{1-\left(\frac{k_1}{k_2}\right)^2\sin^2\theta_1}\\ &= \left[1-\frac{\omega^2\epsilon_1\mu_1}{(\alpha+i\beta)^2}\sin^2\theta_1\right]^{1/2}\\ &\approx 1-\tfrac{1}{2}\,\frac{\omega^2\epsilon_1\mu_1\sin^2\theta_1}{\frac{\omega^2\sigma^2\mu_2^2}{4}(1+i)^2}\\ &\approx 1\end{aligned}$$

The last expression being valid for good conductors. Thus for the transmitted wave we can write

$$\begin{aligned}\mathbf{E}_2 &= \mathbf{E}_{20}\exp\left[i(\mathbf{k}_2.\mathbf{r}-\omega t)\right]\\ &= \mathbf{E}_{20}\exp\left[i\{k_2\cos\theta_2\,x+k_2\sin\theta_2\,z-\omega t\}\right]\\ &= \mathbf{E}_{20}\exp\left[i\{\alpha(1+i)\,x+k_1\sin\theta_1\,z-\omega t\}\right]\\ &\approx \mathbf{E}_{20}\exp(-\alpha x)\exp\left[i\{\alpha x+k_1\sin\theta_1\,z-\omega t\}\right]\end{aligned}$$

For a good conductor $\alpha >> k_1$ and the wave (having an amplitude exponentially decreasing in the x-direction) propagates *along* the x-axis.

PROBLEMS

1. Consider the normal incidence of a plane wave on a dielectric slab as shown in Fig. 20.12. Show that the reflection coefficient is given by

$$R=\frac{2\gamma\,(1+\cos\alpha)}{1+2\gamma\cos\alpha+\gamma^2},$$

where

$$\gamma=\left(\frac{n-1}{n+1}\right)^2,\quad n=\left(\frac{\epsilon}{\epsilon_0}\right)^{1/2},\quad \alpha=\frac{4\pi nt}{\lambda}$$

t being the thickness of the dielectric. Find the values of t for which the reflection coefficient is zero and compare with the results obtained in Sec. 12.2.

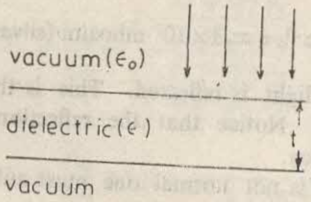


Fig. 20.12 Normal incidence of a plane wave on a dielectric slab.

2. Derive the expression for reflection coefficient for a wave incident normally on a dielectric interface. Notice that the expression is independent of the direction of polarization of the incident beam.

3. Consider a magnetic dielectric with a permeability such that $\frac{\mu}{\mu_0} = \frac{\epsilon}{\epsilon_0}$. Show that for such a material the reflection coefficient for normal incidence is identically equal to zero. This realization is equivalent to the situation where the impedance is matched at the junction of two transmission lines. (The quantity $\sqrt{\frac{\mu}{\epsilon}}$ can be considered as the intrinsic impedance of the medium).
4. A right circularly polarized beam is incident on a perfect conductor at 45°. Show that the reflected beam is left circularly polarized.
5. For polystrene $n \simeq 1.64$. Calculate the Brewster angle. Show that at an angle of incidence of 60°, $r_{\parallel} \simeq -0.0747$, $r_{\perp} \simeq -0.515$. If a right circularly polarized beam is incident at this angle, show that the reflected beam is right elliptically polarized.
6. In the above problem if the angle of incidence is 45°, show that the reflected beam is left elliptically polarized with the ratio of the major and minor axes of the ellipse equal to 3.56. Show also that the refracted beam is right elliptically polarized.
7. Consider a linearly polarized beam with

$$E_x = \tfrac{1}{2}E_0 \cos(\mathbf{k}.\mathbf{r} - \omega t)$$

$$E_y = \frac{1}{\sqrt{2}}\, E_0 \cos(\mathbf{k}.\mathbf{r} - \omega t)$$

$$E_z = -\tfrac{1}{2}E_0 \cos(\mathbf{k}.\mathbf{r} - \omega t)$$

   incident at an angle of 45° on a perfect conductor. Show that the reflected beam is also linearly polarized with its E rotated by 90° (The x, y and z-axes are defined in Fig. 20.1).
8. In the above problem if the beam is incident on glass with $n = 1.5$, determine the state of polarization of the reflected beam.
9. A left circularly polarized beam of intensity 1W/cm^2 is incident on an air-glass interface at 60°. Obtain the states of polarization of the reflected and the refracted beams and the intensities associated with them. Interpret the result physically keeping in mind that the angle of incidence is close to the Brewster's angle.

REFERENCES AND SUGGESTED READING

BRYNGDAHL, O. (1973), Evanescent Waves in Optical Imaging, *Progress in Optics* (Ed. E. Wolf), Vol. XI, North Holland Publishing Co., Amsterdam (Netherlands).

CORSON, D. R. and LORRAIN, P. (1962), *Introduction to Electromagnetic Fields and Waves*, W. H. Freeman and Co., San Francisco (USA).

FEYNMAN, R.P., LEIGHTON, R.B. AND SANDS, M. (1965), *The Feynman Lectures on Physics*, Vol. I, Addison Wesley Publishing Co., Reading, Mass. (USA).

HEIRTZLER, J.R. (1962), The Largest Electromagnetic Waves, *Scient. Amer.*, **206**, (September), 128.

JORDON, E.C. AND BALMAIN, K.G. (1970), *Electromagnetic Waves and Radiating Systems*, Prentice-Hall of India, New Delhi (India).

PANOFSKY, W.K.H. AND PHILLIPS, M. (1962), *Classical Electricity and Magnetism*, Addison Wesley Publishing Co., Reading, Mass. (USA).

REITZ, J.R. AND MILFORD, F.J. (1962), *Foundations of Electromagnetic Theory*, Addison Wesley Publishing Co., Reading, Mass. (USA).

SANDHU, H.S. AND FRIEDMANN, G.B. (1971), Change of Phase on Reflection, *Amer., J. Phys.*, **39**, 388.

SODHA, M.S. and GHATAK, A. K. (1977), *Inhomogeneous Optical Waveguides*, Plenum Publishing Corporation, New York (USA).

PART VI

Photons

21 The Particle Nature of Radiation

> *Two voices are there; one is of the sea,*
> *One of the Mountains; each a mighty voice:*
> *In both from age to age thou didst rejoice,*
> *They were thy chosen music*
>
> WORDSWORTH

21.1 Introduction

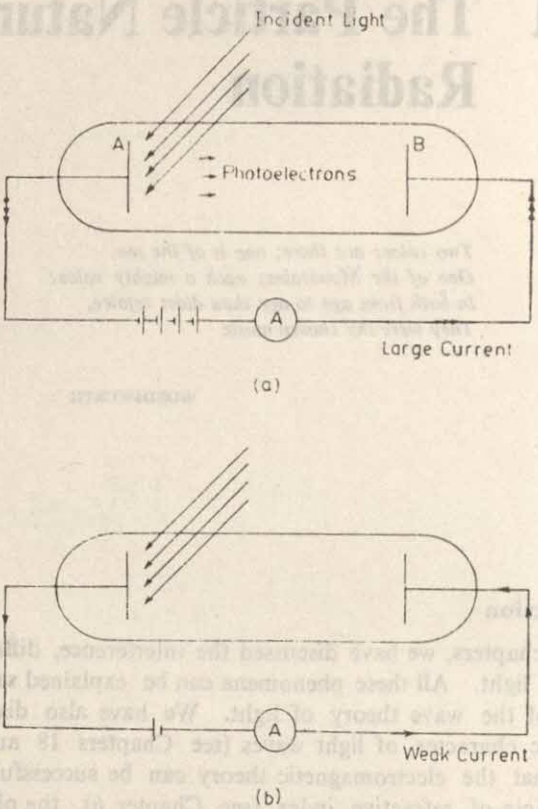
In the earlier chapters, we have discussed the interference, diffraction and polarization of light. All these phenomena can be explained satisfactorily on the basis of the wave theory of light. We have also discussed the electromagnetic character of light waves (see Chapters 18 and 19) and have shown that the electromagnetic theory can be successfully used to explain the origin of refractive index (see Chapter 6), the phenomenon of double refraction (see Chapter 18) and many other experimental results. However, there exist a large number of experimental phenomena which can only be explained on the basis of the corpuscular nature of radiation.* In this chapter we will discuss the famous experiments on the photoelectric effect and the Compton effect which establish the particle nature of light; a wave model is totally inadequate to explain these effects.

21.2 The Photoelectric Effect

As mentioned in Chapter 1, when light falls on certain metal surfaces (like sodium), electrons are emitted. These electrons are known as photoelectrons and can be collected by a metal plate B as shown in Fig. 21.1. The photoelectrons constitute a current between the plates A and B which can be detected by means of an ammeter A. When the voltage across the plates is varied, the current also varies; typical variations of

---

*In the first chapter, we discussed how one can reconcile to the dual nature of radiation (i.e. the wave and the particle aspects) on the basis of the quantum theory.



(a) (b)

Fig. 21.1 If light (of certain frequency) is allowed to fall on a metal like sodium, electrons are emitted which can be collected by the plate B. (a) and (b) correspond to positive and negative voltage applied to the plate B. Even when the plate is kept at a low negative voltage, one can detect a small current.

the current with voltage are shown in Fig. 21.2. The figure corresponds to monochromatic light of a particular wavelength and the different curves correspond to different intensities of the beam. From the figure one can draw the following conclusions:

(i) At zero voltage there is a finite value of the current implying that some of the emitted photoelectrons reach the metal surface B.

(ii) As the voltage is increased, the current increases till it reaches a saturation value; this will happen when all the emitted photoelectrons are collected by the plate B.

(iii) If the plate B is kept at a slightly negative potential, there is a weak current implying that some of the photoelectrons do manage to reach the plate B. However, beyond a certain voltage (which is shown as $-V_c$ in the figure) the current is zero; V_c is known as the cutoff voltage and the quantity $|e| V_c$ will represent the maximum kinetic energy of the photoelectrons (e represents the charge of the electron).

(iv) If we do not change the wavelength of the incident radiation but make it more intense, the magnitude of the current will become larger as shown in Fig. 21.2 implying a greater emission of the photoelectrons. Notice that the value of the cutoff potential remains the same; this important result implies that the maximum kinetic energy of the emitted photoelectrons does not depend on the intensity of the incident radiation.

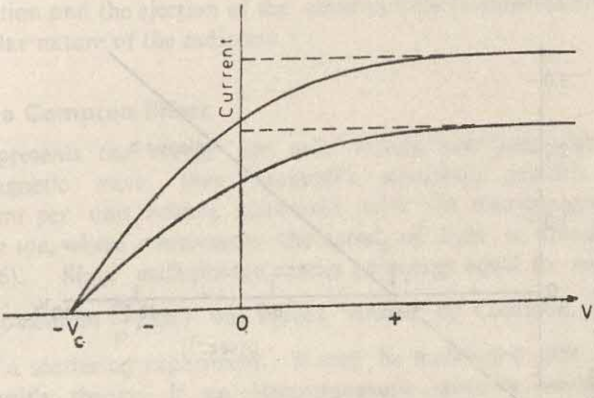


Fig. 21.2 Typical variation of the photocurrent with voltage. The upper curve corresponds to light (of the same frequency) having greater intensity.

(v) If the frequency of the incident radiation is increased then the cutoff potential (and hence the maximum kinetic energy of the emitted photoelectrons) increases. Indeed the maximum kinetic energy of the electrons ($=|e| V_c$) varies linearly with the frequency as shown in Fig. 21.3. Further, for frequencies less than a critical value (shown as ν_c in Fig. 21.3), there is *no* emission of photoelectrons no matter what the intensity of the incident radiation may be.

At first sight it appears that since electromagnetic waves carry energy (see Chapter 19), the wave model for light should be able to explain the emission of photoelectrons from a metal surface. However, there are certain peculiarities associated with photoelectric effect which cannot be satisfactorily explained by means of a wave model. The first peculiarity is the fact that the maximum kinetic energy of the electrons does not depend on the intensity of the incident radiation, it only depends on its frequency; further, a greater intensity leads to a larger number of electrons constituting a larger current. Thus, a faint violet light would eject electrons of greater kinetic energy than an intense yellow light although the latter would produce a large number of electrons. A wave model would, however, predict that a large intensity of the incident radiation would result in a greater kinetic energy of the emitted electrons. Indeed, as mentioned earlier, the observed maximum kinetic energy of the photoelectrons is linearly related to the frequency of the incident radiation and one may write (see Fig. 21.3)

$$T_{\max} = -B + h\nu = h(\nu - \nu_c) \tag{21.2-1}$$

where $B\ (= h\nu_c)$ is a constant and h is the Planck's constant $(= 6.627 \times 10^{-27}$ erg.-sec). The frequency ν_c represents the cutoff frequency and is a characteristic of the metal. For example, for sodium the cutoff frequency is $4.39 \times 10^{14}\ \text{sec}^{-1}$.

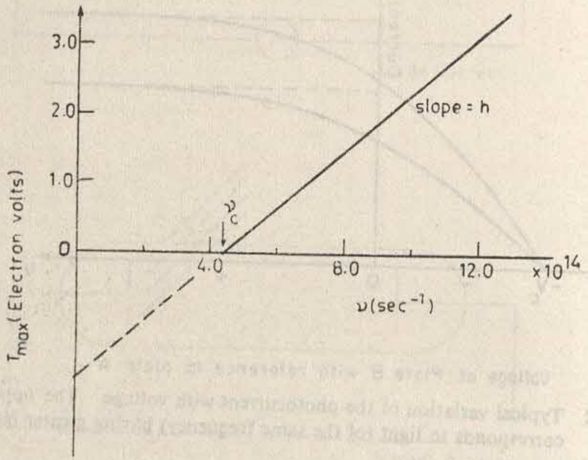


Fig. 21.3 The variation of the maximum kinetic energy of the electrons as a function of frequency of the incident light.

The second peculiarity is the fact that there is almost no time lag between the times of incidence of the radiation and the ejection of the photoelectron. For weak intensities of the incident beam the wave theory predicts a considerable time lag for the electrons to absorb enough energy to leave the metal surface. This can be illustrated by considering a specific example. One can observe a detectable photocurrent if the surface of sodium metal is illuminated by violet light of intensity as low as 10^{-10} W/sq. cm. Assuming that the energy is uniformly absorbed by the upper ten layers of sodium, each atom would receive energy at the rate of $\sim 10^{-26}$ W* ($\approx 10^{-7}$ eV/sec). Assuming that an electron should acquire an energy $\sim$ 1eV to escape from the metal, one should expect a time lag of order 10^7 sec. ($\approx$ few months).

*Ten layers of sodium will contain

$$\frac{6 \times 10^{23} \times 10 \times 10^{-8}}{23} \approx 2 \times 10^{15} \text{ atoms/sq. cm,}$$

where we have assumed the density of sodium to be $\approx 1\text{g/cm}^3$.

However, the experiments show that there is no detectable time lag between the incidence of the radiation and the emission of the photoelectrons.*

Einstein's *photon* theory gives a very satisfactory explanation of the photoelectric effect. According to this theory, a light beam (of frequency ν) essentially consists of individual corpuscles called photons. Each photon carries an energy equal to $h\nu$. This corpuscular model can explain all the observations discussed above. Thus, for all frequencies below the cutoff ν_c, each photon will carry energy less then $h\nu_c$ which will not be sufficient to eject the electron from the metal. For $\nu > \nu_c$, a major fraction of the excess energy $[= h(\nu - \nu_c)]$ appears as kinetic energy of the emitted electron. Further, the non-measurable time lag between the incidence of the radiation and the ejection of the electron follows immediately from the corpuscular nature of the radiation.

21.3 The Compton Effect

If u represents the energy per unit volume associated with a plane electromagnetic wave, then Maxwell's equations predict that the momentum per unit volume associated with the electromagnetic wave would be u/c, where c represents the speed of light in free space (see Sec. 19.6). Since each photon carries an energy equal to $h\nu$, it should have a momentum $\frac{h\nu}{c}$; this was indeed verified by Compton in 1923 by means of a scattering experiment. It may be mentioned that according to Maxwell's theory, if an electromagnetic wave is incident on an electron (which may be bound to an atom), then the electron executes oscillatory motion of the same frequency as of the electromagnetic wave. The oscillating electron radiates electromagnetic waves (of the *same* frequency) in all directions; indeed this scattering leads to the blue colour of the sky and also to the redness of the setting sun.

On the other hand, the quantum theory tells us that since a certain velocity is imparted to the electron, the scattered photon must have a smaller energy and hence a larger wavelength. The change in the wavelength can easily be calculated in the following manner. Consider the scattering of a photon by an electron as shown in Fig. 21.4. The scattered photon is assumed to have a frequency ν'. Conservation of energy leads to

$$h\nu = h\nu' + E_k \tag{21.3-1}$$

where E_k represents the kinetic energy imparted to the electron. Conserving the x-and y-components of the momentum, we have

$$\frac{h\nu}{c} = \frac{h\nu'}{c}\cos\theta + p\cos\varphi \tag{21.3-2}$$

---

*Indeed, Lawrence and Beams (1928) had devised an experiment to find out whether the time lag was $\lesssim 3\times10^{-9}$ sec; the experiment gave a negative result.

and

$$0 = \frac{h\nu'}{c}\sin\theta - p\sin\varphi \qquad (21.3\text{-}3)$$

where p represents the momentum of the electron after collision, θ and φ represent the angles made by the scattered photon and the electron with the original direction of the photon (see Fig. 21.4). It will be shown

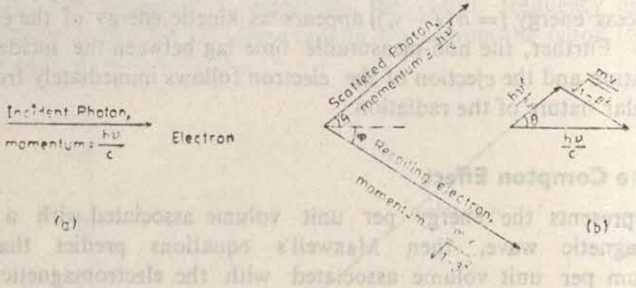

Incident Photon,
momentum $= \frac{h\nu}{c}$
Electron
(a)
Scattered Photon, momentum $= \frac{h\nu'}{c}$
θ
φ
Recoiling electron, momentum $= \frac{m_0 v}{\sqrt{1-\beta^2}}$
$\frac{h\nu'}{c}$
$\frac{m_0 v}{\sqrt{1-\beta^2}}$
θ
$\frac{h\nu}{c}$
(b)

Fig. 21.4 The Compton scattering of a photon by an electron.

that for a measurable Compton effect, the frequency ν should be in the X-ray or in the γ-ray region.* For such high energy photons, the velocity imparted to the electron is comparable to the speed of light and one must use proper relativistic expressions for E_k and p. According to the theory of relativity, the kinetic energy E_k would be given by

$$E_k = E - m_0c^2 = mc^2 - m_0c^2 = \frac{m_0c^2}{\sqrt{1-\beta^2}} - m_0c^2, \qquad (21.3\text{-}4)$$

where $\beta = \frac{v}{c}$, m_0 represents the rest mass of the electron, v the speed of the electron and c the speed of light in free-space; the quantities E and m_0c^2 are known as the total energy and the rest-mass energy of the electron. Further, the relativistic momentum of the electron is given by

$$p = mv = \frac{m_0 v}{\sqrt{1-\beta^2}} \qquad (21.3\text{-}5)$$

Now,

$$p^2c^2 + m_0^2c^4 = \frac{m_0v^2}{1 - v^2/c^2} + m_0^2c^4$$

$$= \frac{m_0^2c^4}{1 - v^2/c^2} \qquad (21.3\text{-}6)$$

---

*For X-rays, $\lambda \lesssim 1$ Å and $h\nu \gtrsim 10^4$ eV.

or

$$p^2c^2 + m_0^2c^4 = E^2 = (E_k + m_0c^2)^2 = E_k^2 + m_0^2c^4 + 2E_km_0c^2 \qquad (21.3\text{-}7)^*$$

Thus

$$E_k^2 + 2E_k m_0c^2 = p^2c^2 \qquad (21.3\text{-}8)$$

Substituting for E_k from Eq. (21.3-1), we get

$$h^2(\nu - \nu')^2 + 2h(\nu - \nu')\, m_0c^2 = p^2c^2 \qquad (21.3\text{-}9)$$

Further, Eqs. (21.3-2) and (21.3-3) can be rewritten in the form

$$p \cos\varphi = \frac{h\nu}{c} - \frac{h\nu'}{c}\cos\theta \qquad (21.3\text{-}10)$$

and

$$p \sin\varphi = \frac{h\nu'}{c}\sin\theta \qquad (21.3\text{-}11)$$

In order to eliminate φ, we square and add to obtain

$$p^2 = \left(\frac{h\nu}{c}\right)^2 + \left(\frac{h\nu'}{c}\right)^2 - \frac{2h^2\nu\nu'}{c}\cos\theta \qquad (21.3\text{-}12)$$

Substituting in Eq. (21.3-9), we obtain

$$h^2(\nu^2 - 2\nu\nu' + \nu'^2) + 2h(\nu - \nu')\, m_0c^2 = h^2\nu^2 + h^2\nu'^2 - 2h^2\nu\nu'\cos\theta$$

or

$$\frac{2h(\nu - \nu')\, m_0c^2}{2\nu\nu'} = h^2(1 - \cos\theta)$$

or

$$\Delta\lambda \equiv \lambda' - \lambda = \frac{h}{m_0c}(1 - \cos\theta) \qquad (21.3\text{-}13)$$

which gives us the Compton shift.** If we substitute the values of h, m_0 and c, we obtain

$$\Delta\lambda = \lambda' - \lambda = 0.0485 \sin^2\theta/2 \qquad (21.3\text{-}14)$$

where λ is measured in Angstroms. The above equation shows that the maximum change in the wavelength is about 0.05 Å, and as such for a measurable shift one must use radiation of small wavelength. In Fig. 21.5 we have shown the wavelength of the scattered photon at different angles with respect to the primary beam as obtained by Compton in his original experiment* in 1923. The solid curve corresponds to Eq. (21.3-14) with $\lambda = 0.022$ Å. (Notice that the corresponding photon energy is $\sim 6.6\times10^{-27}\times3\times10^{10}/2.2\times10^{-10}$ ergs $\simeq 0.5$ MeV, which corresponds to a γ-ray). The good agreement between theory and experiment proves that radiation behaves as if it consists of corpuscles of energy $h\nu$ having a momentum $h\nu/c$.

---

*The equation $E^2 = m_0^2c^4 + p^2c^2$, is valid for all particles. For the photon, $E = h\nu = pc$ giving a zero value for the rest mass.

**In the derivation of the Compton shift we have assumed that the electron is free although we know that the electrons are bound to the atoms. The assumption of a free electron is justified because the binding energy ($\approx$ few eV) is usually very much smaller in comparison to the photon energy ($\gtrsim$ 1000 eV).

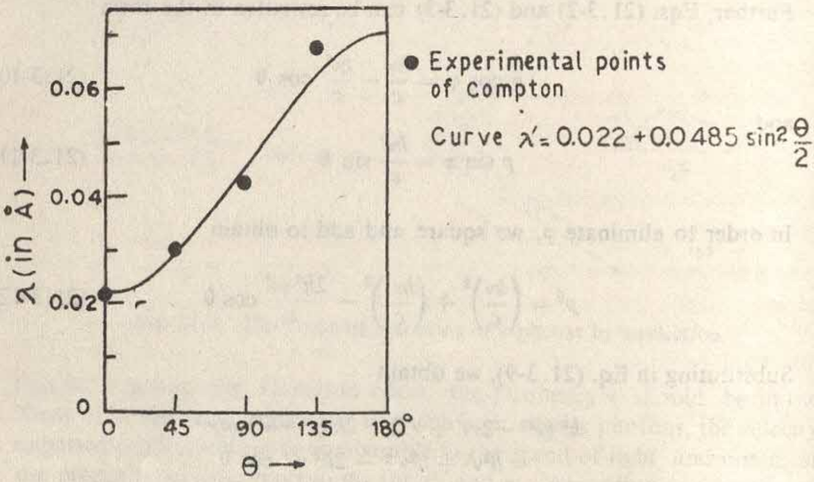


Fig. 21.5 The variation of wavelength of the scattered photon with the angle of scattering. The solid curve corresponds to Eq. (21.3-14) with $\lambda = 0.022$ Å. The dots represent the experimental points obtained by Compton. The figure has been adapted from the original paper of A.H. Compton on "A Quantum Theory of the Scattering of X-rays by light Elements", *Physical Review*, 1923, **21**, 382.

In Fig. 21.6 we have given the schematic of the experimental arrangement for the measurement of the Compton shift. A monochromatic beam of X-rays (or γ-rays) is allowed to fall on a sample scatterer and the scattered photons were detected by means of a crystal spectrometer. The crystal spectrometer allows one to find the intensity distribution (as a function of λ) for a given value of θ. The experimental findings of Compton are shown in Fig. 21.7; the experiment corresponds to the Molybdenum K_α line ($\lambda = 0.711$ Å). The sample used was graphite. Notice that at each value of θ, there are two peaks; the first peak appears at almost the same wavelength as the primary beam. This peak is because of the fact that the photon may be scattered by the whole atom; consequently, the

---

* The original papers of Compton have been published in a book on *Scientific Papers of A.H. Compton: X-ray and Other Studies*, edited by R.S. Shankland, The University of Chicago Press, 1973.

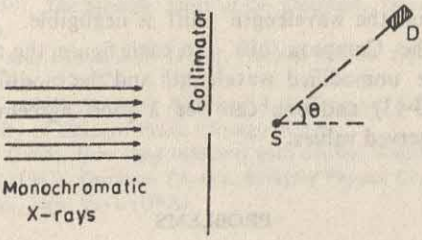


Fig. 21.6 Outline of the experimental arrangement for the measurement of the Compton shift. A collimated beam of monochromatic X-rays is scattered by the scatterer S; the wavelength of the scattered photon is measured by the detector D.

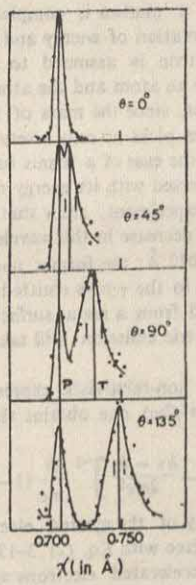


Fig. 21.7 The intensity variation as a function of the wavelength of the scattered photon. The vertical line (marked P) corresponds to the unmodified wavelength $\lambda = 0.711$ Å. The second vertical line (marked T) corresponds to the wavelength as predicted by Eq. (21.3–13). The figure has been adapted from A.H. Compton's paper on "The spectrum of Scattered X-rays", *Physical Review*, 1923, 22, 409.

quantity m_0 appearing in Eq. (21.3-13) is not the electron mass but the mass of carbon atom (which is about 22,000 times that of the mass of the electron). Thus the wavelength shift is negligible. The second peak corresponds to the Compton shift. In each figure, the two vertical lines correspond to the unmodified wavelength and the modified wavelength as given by Eq. (21.3-13) and one can see a good agreement between the predicted and observed values.

PROBLEMS

1. (a) Calculate the number of photons emitted per second by a 5 mW laser assuming that it emits light of wavelength 6328 Å: (Ans. 1.6×10^{16}).
   (b) The beam is allowed to fall normally on a plane mirror. Calculate the force acting on the mirror. (Ans. 3.3×10^{-11} N).
2. Assume a 40 W sodium lamp ($\lambda \approx 5893$ A) emitting light in all directions. Calculate the rate at which the photons cross an unit area placed normally to the beam at a distance of 10 m from the source. (Ans. $\approx 10^{17}$ photons/m^2-sec).
3. In the photoelectric effect, a photon is completely absorbed by the electron. Show that the laws of conservation of energy and momentum cannot be satisfied simultaneously if a free electron is assumed to absorb the photon. (Thus the electron has to be bound to an atom and the atom undergoes a recoil when the electron is ejected. However, since the mass of the atom is much larger than that of the electron, the atom picks up only a very small fraction of the energy, this is somewhat similar to the case of a tennis ball hitting a heavy object; the momentum of the ball is reversed with its energy remaining almost the same).
4. In the Compton scattering experiment, show that the fractional loss of energy of the photon increases with decrease in the wavelength. Calculate this fractional loss for $\lambda = 0.711$ Å and 0.022 Å; the former corresponds to the Molybdenum K_α X-ray line and the latter to the γ-rays emitted from RaC.
5. If photoelectrons are emitted from a metal surface by using blue light, can you say for sure that photoelectric emission will take place with yellow light and with violet light?
6. Show that if one used the non-relativistic expression for the kinetic energy of the electron (namely, $E_k = p^2/2m$), one obtains the following expression for the Compton shift:

   $$\lambda' - \lambda = \left[1 - \frac{h\nu - h\nu'}{2mc^2}\right]^{-1} \frac{h}{mc}(1 - \cos\theta)$$

   Show that when the velocity of the ejected electron is small compared to c, the above equation would agree with Eq. (21.3–13).
7. X-rays are emitted when accelerated electrons are stopped by a target. In a particular experiment, one finds that the minimum wavelength of the X-rays produced by 20 keV electrons is about 0.6 Å. Estimate the value of the Planck's constant.

REFERENCES AND SUGGESTED READING

Born, M. (1951), *The Restless Universe*, Dover Publications, New York (USA).
Compton, A.H. (1923), A Quantum theory of the scattering of X-rays by light elements, *Phys. Rev.*, **21**, 382

DE-BROGLIE, L. (1937), *Matter and Light: The New Physics*, Dover Publications, New York (USA).

HOFFMAN, B. (1959), *The Strange Story of the Quantum*, Dover Publications, New York (USA).

LAWRENCE, E.O. AND BEAMS, J.W. (1928), The element of time in the photoelectric effect, *Phys. Rev.*, 32, 478.

SHANKLAND, R.S. (1973), *Scientific Papers of A.H. Compton: X-Ray and Other Studies*, The University of Chicago Press, Chicago (USA).

WEISSKOPF, V.W. (1968), How light interacts with matter, *Scient. Amer.* (September).

WICHMANN, E.H. (1967), *Quantum Physics*, Berkeley Physics Course, Vol. 4, McGraw-Hill Book Co., New York (USA).

PART VII

Sources of Coherent Light

22 Lasers: An Elementary Account

When you turn your eyes to heaven
skyward to the azure flow,
when at dusk the Sun is driven
down in crimson fireglow
There in Nature's deepest kernel
healthy, glad of heart and sight
you perceive the great eternal
essence of chromatic light.

Goethe on Newton's experiment
(*Translation by* DOUGLAS WORTH & VICTOR F. WEISSKOPF)

22.1 Introduction

The light emitted from a conventional light source (like a sodium lamp) is said to be incoherent because the radiation emitted from different atoms do not, in general, bear any definite phase relationship with each other. On the other hand, the light emitted from a laser* has a very high degree of coherence (see Chapter 14) and is almost perfectly parallel. Thus, if the laser beam is focussed by a convex lens, it is focussed in a region whose linear dimension is $\sim \frac{\lambda f}{a}$, where λ is the wavelength of the laser beam, f the focal length of the lens and a, the aperture of the beam. If $f \sim a$, the area of the region to which the laser beam can be focussed is $\sim \lambda^2$ and because $\lambda \sim 5 \times 10^{-5}$ cm, one obtains a tremendous intensity in this region.** This high intensity can be used to effectively 'drill' through a metallic target.

*Laser is an acronym for Light Amplification by Stimulated Emission of Radiation. As the name implies, the process of stimulated emission is employed in the amplification of a light beam (see Sections 22.2 and 22.3). The theory of a laser was proposed in 1958 by Townes and Schawlow in the USA and by Basov and Prokhorov in the USSR. [Townes, Basov and Prokhorov were awarded the 1964 Nobel Prize for the invention of the Maser and the Laser; Maser is an acronym for Microwave Amplification by Stimulated Emission of Radiation.] The first laser was fabricated by Maiman in 1960.

**For example, if the power associated with the laser beam ($\lambda \approx 5 \times 10^{-7}$ m) is 10 W, then the intensity at the focal plane would be $\sim \frac{10}{25 \times 10^{-14}} = 4 \times 10^{15}$ W/m^2!

Because of the high degree of coherence associated with a laser beam, it finds important applications in many diverse areas like holography, data processing, communications, etc. [see, e.g., Schawlow (1969), Sobolev (1974), Lengyl (1966), Siegman (1971), Sodha and Ghatak (1977)].

The basic physics involved in the lasing action is the phenomenon of stimulated emission which will be discussed in Sec. 22.2. In this emission, the electromagnetic radiation forces an atom in the excited state to emit radiation which has the same frequency, same phase, and which propagates in the same direction as the incident radiation (Fig. 22.1). These can cause further stimulated emissions and this process leads to a coherent radiation. However, for light amplification to take place in the system, one must have a larger number of atoms in the higher energy state than in the lower energy state (Sec. 22.3). This is known as population inversion. In the first laser fabricated by Maiman in 1960, population inversion was achieved in the following manner: The laser was made from a single cylindrical crystal of ruby whose ends were flat and one of the ends was completely silvered and the other end was partially silvered (Fig. 22.2). Ruby consists of Al_2O_3 with some of the aluminum atoms replaced by chromium.* The energy states of the chromium ion are

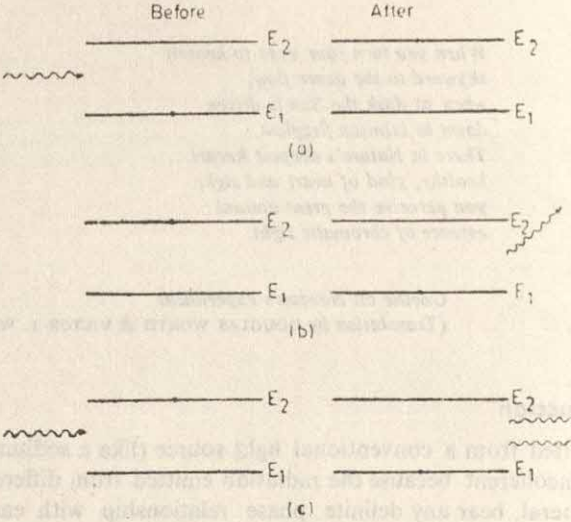


Fig. 22.1 **(a) Absorption of a photon which makes the atom to go to an excited state. (b) The spontaneous emission. (c) The stimulated emission.**

---

*The Al_2O_3 crystal which serves as a medium to suspend the chromium ions is known as the host crystal. The characteristics of the host crystal affect the laser action and also the broadening of the energy levels of the activator atoms which in this case are that of chromium. For a good lasing action, the ruby crystal consists of about 0.05% of chromium; however, higher concentrations of chromium have also been used. For a detailed discussion about host crystals, see for example Brown (1969).

shown in Fig. 22.3. The chief characteristic of the energy levels of a chromium ion is the fact that the bands labelled as E_1 and E_2 have a lifetime* which is $\lesssim 10^{-8}$ sec whereas the state marked M has a lifetime which is $\sim 3 \times 10^{-3}$ sec. A state which is characterised by such a long life is termed as a metastable state.

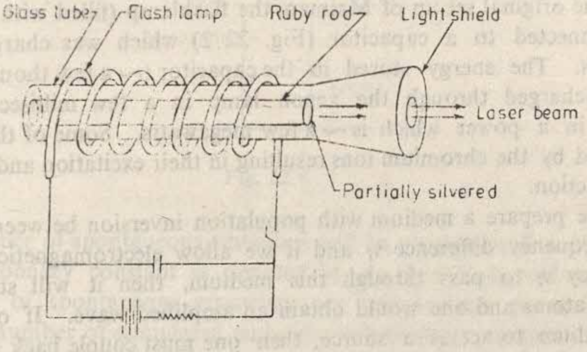


Fig. 22.2 **The ruby laser.**

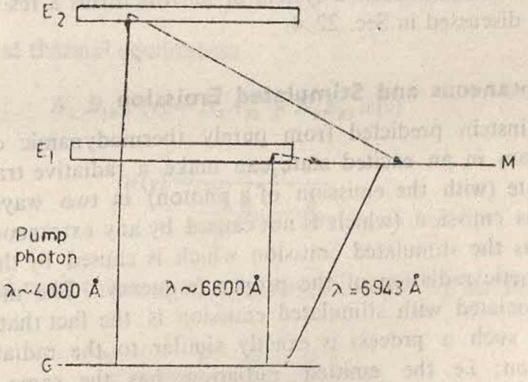


Fig. 22.3 The energy levels of the chromium ion. G represents the ground state and M the metastable state.

The chromium ion in its ground state can absorb a photon (whose wavelength is around 6600 Å) and make a transition** to one of the states in the band E_1; it could also absorb a photon of $\lambda \sim 4000$ Å and make a transition to one of the states in the band E_2. In either case, it immediately makes a non-radiative transition* (in a time $\lesssim 10^{-9}$ sec) to the metastable state M and since the state M has a very long life, the number of atoms in this state keep on increasing and one may achieve a population inversion between the states M and G. Thus we may have a larger number of atoms in the level M to that in the state G. Once population inversion is achieved, light amplification can take place.

---

*The lifetime represents the average time that an atom spends in an excited state before making a transition to a lower energy state.

**This is known as optical pumping and the photons which are absorbed by the chromium ions are produced by the flash lamp (Fig. 22.2).

In the original set up of Maiman, the flashlamp (filled with xenon gas) was connected to a capacitor (Fig. 22.2) which was charged to a few kilovolts. The energy stored in the capacitor ($\sim$ a few thousand joules) was discharged through the xenon lamp in a few milliseconds. This results in a power which is $\sim$ a few megawatts. Some of this energy is absorbed by the chromium ions resulting in their excitation and subsequent lasing action.

If we prepare a medium with population inversion between two levels with frequency difference ν_l and if we allow electromagnetic energy of frequency ν_l to pass through this medium, then it will stimulate the excited atoms and one would obtain an amplified wave. If one requires the medium to act as a source, then one must couple back some of the output energy back into the medium and at the same time prepare the atoms in a state of population inversion. The coupling back of energy into the medium is carried out through a system of mirrors placed at the ends of the medium. Such a system of mirrors forms a resonant cavity** and will be discussed in Sec. 22.4.

22.2 Spontaneous and Stimulated Emission

In 1917, Einstein predicted from purely thermodynamic considerations that an atom in an excited state can make a radiative transition to the ground state (with the emission of a photon) in two ways; one is the spontaneous emission (which is not caused by any extraneous effects) and the other is the stimulated emission which is caused by the presence of electromagnetic radiation of the proper frequency. The most important feature associated with stimulated emission is the fact that the radiation emitted in such a process is exactly similar to the radiation inducing the transition; i.e. the emitted radiation has the same phase, same polarization, same frequency as the incident radiation and propagates in the same direction.

Consider an atom having two states and let N_1 and N_2 be the number of atoms per unit volume in states 1 and 2 respectively (see Fig. 22.4). Let $u(\nu)$ represent the energy density of radiation; i.e. $u(\nu)\, d\nu$ represents the energy density in the frequency interval between ν and $\nu + d\nu$. Now.

---

*In a non-radiative transition, the excess energy is absorbed by the lattice and does not appear in the form of electromagnetic radiation.

**In the ruby laser, the two silvered ends of the ruby rod act as mirrors and form the resonant cavity.

the rate of absorption of radiation of frequency ν (which will lead to the transition $1 \to 2$) is assumed to be proportional to the radiation energy density $u(\nu)$. Thus, if the proportionality constant is denoted by B_{12} then $N_1 B_{12} u(\nu)$ will represent the number of absorptions per unit volume per unit time.

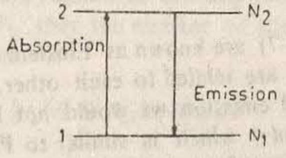


Fig. 22.4

The number of spontaneous emissions will be proportional to N_2 and if the proportionality constant is denoted by A_{21} then $N_2 A_{21}$ will represent the number of spontaneous emissions per unit volume per unit time. Finally, the number of stimulated emissions will be proportional to N_2 and also to the radiation energy density $u(\nu)$. If the proportionality constant is denoted by B_{21} then $N_2 B_{21} u(\nu)$ will represent the number of stimulated emissions per unit volume per unit time. The quantities A_{21}, B_{12} and B_{21} are known as Einstein coefficients and are determined by the atomic system.

Obviously, at thermal equilibrium

$$N_1 B_{12} u(\nu) = N_2 A_{21} + N_2 B_{21} u(\nu) \tag{22.2-1}$$

or

$$u(\nu) = \frac{A_{21}}{\frac{N_1}{N_2} B_{12} - B_{21}} \tag{22.2-2}$$

Now, at thermal equilibrium N_1 and N_2 will be related by the Boltzmann's law

$$\frac{N_2}{N_1} = \exp\left[-\frac{E_2 - E_1}{k_B T}\right] \tag{22.2-3}$$

where k_B is the Boltzmann's constant and T represents the temperature of the system. Substituting in Eq. (22.2-1), we get

$$u(\nu) = \frac{A_{21}}{e^{h\nu/k_B T} B_{12} - B_{21}} \tag{22.2-4}$$

On the other hand, Planck's law tells us that

$$u(\nu) = \frac{8\pi h\nu^3}{c^3} \frac{1}{e^{h\nu/k_B T} - 1} \tag{22.2-5}$$

Comparing Eqs. (22.2-4) and (22.2-5) we obtain*

$$B_{12} = B_{21} \tag{22.2-6}$$

and

$$\frac{8\pi h\nu^3}{c^3} B_{21} = A_{21} \tag{22.2-7}$$

Eqs. (22.2-6) and (22.2-7) are known as Einstein relations and show that the Einstein coefficients are related to each other. Notice that if we had not assumed stimulated emission we would not have been able to arrive at an expression for $u(\nu)$ which is similar to Planck's law. Indeed, in order to obtain the correct form for $u(\nu)$, Einstein predicted the existence of stimulated emission which was later confirmed by rigorous quantum theory [see, for example, Ghatak and Lokanathan (1975)].

22.3 Population Inversion

Consider the propagation of a collimated beam of light along the x-direction. We wish to calculate the change in intensity of the beam as it propagates through a layer of atoms which are bounded by two planes (normal to the x-axis; see Fig. 22.5) and are separated by a distance dx.

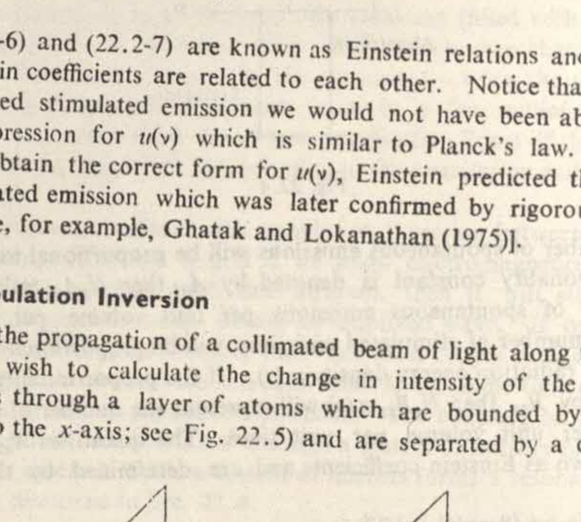


Fig. 22.5

If the cross-sectional area of each of the planes is S then the volume of the layer would be Sdx.

If $n_1(\nu)\, d\nu$ represents the number of atoms per unit volume** which are in level 1 and are capable of absorbing radiation whose frequency lies between ν and $\nu + d\nu$ then the number of upward transitions ($1 \rightarrow 2$) per unit time, in the layer of volume Sdx, would be

$$[n_1(\nu)\, d\nu\, B_{12}\, u(\nu)]\, S\, dx$$

---

*In the presence of degeneracies Eq. (22.2-7) would have remained the same. However, Eq. (22.2-6) would have modified to $\frac{g_1}{g_2} B_{12} = B_{21}$, where g_1 and g_2 represent the degree of degeneracy of levels 1 and 2 respectively.

**$\int n_1(\nu)\, d\nu = N_1$

In each transition a photon of energy $h\nu$ is absorbed. Thus, the energy lost per unit time from the radiation would be

$$h\nu \, [n_1(\nu) \, d\nu \, B_{12} \, u(\nu)] \, S \, dx$$

Similarly, if $n_2(\nu) \, d\nu$ represents the number of atoms per unit volume* which are in level 2 and are capable of emitting radiation whose frequency lies between ν and $\nu + d\nu$, then the number of stimulated emissions per unit time in the layer would be

$$[n_2(\nu) \, d\nu \, B_{21} \, u(\nu)] \, S \, dx$$

In each transition a photon of energy $h\nu$ is emitted and this reinforces the propagating beam. Thus, the energy gained per unit time would be

$$h\nu \, [n_2(\nu) \, d\nu \, B_{21} \, u(\nu)] \, S dx$$

[We have neglected the effect of spontaneous emissions; this is because a photon emitted in a spontaneous process is in a random direction and as such does not, in general, contribute to the beam (see Fig. 22.1)].

Since $u(\nu) \, d\nu \, S \, dx$ represents the energy contained in the layer, in the frequency range ν and $\nu + d\nu$, we may write

$$\left[\frac{\partial}{\partial t} u(\nu) \, S \, d\nu \, dx \right] = h\nu \, [-n_1(\nu) \, d\nu \, B_{12} \, u(\nu) + n_2(\nu) \, d\nu \, B_{21} \, u(\nu)] \, S \, dx$$

or

$$\frac{\partial}{\partial t} u(\nu) = -h\nu \left[B_{12} \, n_1(\nu) - B_{21} \, n_2(\nu) \right] u(\nu) \qquad (22.3\text{-}1)$$

Now, if I_ν represents the intensity of the beam; i.e. if $I_\nu d\nu$ represents the energy crossing an unit area** per unit time whose frequencies lie between ν and $\nu + d\nu$ then

$$[I_\nu(x + dx) \, d\nu - I_\nu(x) \, d\nu] \, S$$

will represent the rate at which the energy is flowing *out* of the layer. Since $u(\nu) \, d\nu \, S \, dx$ represents the radiation energy contained in the layer whose frequency lies between ν and $\nu + d\nu$, we will have

$$[I_\nu(x + dx) \, d\nu - I_\nu(x) \, d\nu] \, S = + \frac{\partial}{\partial t} [u(\nu) \, d\nu \, S dx]$$

or

$$\frac{\partial u(\nu)}{\partial t} = \frac{I_\nu(x + dx) - I_\nu(x)}{dx} = \frac{\partial I_\nu}{\partial x} \qquad (22.3.2)$$

---

*$\int n_2(\nu) \, d\nu = N_2$

**In this case perpendicular to the x-axis.

Using Eq. (22.3-1) we obtain

$$\frac{\partial I_\nu}{\partial x} = -h\nu\,[B_{12}n_1(\nu) - B_{21}n_2(\nu)]\,u(\nu) \tag{22.3-3}$$

But the intensity of the beam is related to the energy density through the following relation

$$u(\nu)\,\mathrm{v} = I_\nu \tag{22.3-4}$$

where $\mathrm{v}\left(=\frac{c}{\eta}\right)$ represents the speed of propagation of the wave; η being the refractive index. Substituting the above expression for $u(\nu)$ in Eq. (22.3-3), we get

$$\frac{\partial}{\partial x} I_\nu = -\frac{h\nu\, B}{\mathrm{v}}\,[n_1(\nu) - n_2(\nu)]\, I_\nu$$

where $B = B_{12} = B_{21}$ (see Sec. 22.2). Thus

$$\frac{1}{I_\nu}\frac{\partial I_\nu}{\partial x} = -\frac{h\nu\eta}{c}\,[n_1(\nu) - n_2(\nu)]\,B \tag{22.3-5}$$

Now, if a beam is propagating in an absorbing medium, the loss of intensity, $-dI_\nu$, due to absorption would be proportional to I_ν and to dx; hence we may write

$$dI_\nu = -\alpha_\nu I_\nu\, dx \tag{22.3-6}$$

where α_ν represents the proportionality constant and is known as the absorption coefficient. Thus

$$\frac{dI_\nu}{dx} = -\alpha_\nu I_\nu \tag{22.3-7}$$

or

$$I_\nu(x) = I_\nu(x=0)\,e^{-\alpha_\nu x} \tag{22.3-8}$$

If we compare Eq. (22.3-5) with Eq. (22.3-7), we may write

$$\alpha_\nu = \frac{h\nu\eta}{c}\,[n_1(\nu) - n_2(\nu)]\,B \tag{22.3-9}$$

Since $n_1(\nu)$ and $n_2(\nu)$ are very sharply peaked around $\nu = \nu_0$, we may integrate Eq. (22.3-9) to obtain*

$$\begin{aligned}\int \alpha_\nu\, d\nu &= \frac{h\nu_0\eta}{c}\,[N_1 - N_2]\,B \\ &= \kappa\,(N_1 - N_2)\end{aligned} \tag{22.3-10}$$

---

*In the presence of degeneracies one obtains

$$\int \alpha_\nu\, d\nu = \frac{c^2 A}{8\pi\nu_0^2\,\eta^2}\,\frac{g_2}{g_1}\left(N_1 - \frac{g_1}{g_2}N_2\right)$$

where

$$\kappa = \frac{h\nu_0\eta}{c} B = \frac{h\nu_0\eta}{c} \frac{c^3}{8\pi h\nu_0^3\eta^3} A$$
$$= \frac{c^2}{8\pi\eta^2\nu_0^2} A \qquad (22.3\text{-}11)$$

represents the integrated absorption cross-section per atom.* Eq. (22.3-10) is known as the Fuchtbauer-Landenburg formula.

As it is obvious from Eq. (22.3-9), α_ν must be a very sharply peaked function around the absorption line**; and a typical plot is given in Fig. 22.6. Thus we may write

$$\alpha_\nu = \kappa (N_1 - N_2)\, g(\nu)$$

where the function $g(\nu)$ is sharply peaked around $\nu = \nu_0$ and

$$\int g(\nu)\, d\nu = 1$$

At thermal equilibrium $N_1 > N_2$ (i.e. the population of the ground

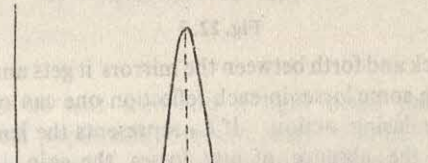


Fig. 22.6 Typical variation of α_ν with ν.

*When a beam of moderate intensity is propagating through a material, only a very small fraction of atoms are excited to the higher state and as such $\frac{N_2}{N_1} << 1$; under this approximation we obtain

$$\int \alpha_\nu\, d\nu = \kappa N_0$$

where N_0 represents the total number of atoms/unit volume.

**The width $\Delta\nu$ is, in general, due to a variety of phenomena such as thermal motion, collision of atoms, etc. In particular, for thermal motion

$$\Delta\nu \simeq \frac{2\nu_0}{c}\left[\frac{2kT\, ln\, 2}{m}\right]^{1/2}$$

(*Contd.*)

state is greater than the population of the excited state) and as can be seen from Eq. (22.3-9), α_ν is positive which implies absorption of the beam as it propagates through the material, physically, the energy is used up in the excitation of the atoms to a higher energy state.

On the other hand, if we have a situation in which $N_2 > N_1$, then α_ν would be negative and the beam would get amplified; this process is referred to as light amplification. Since this corresponds to a higher population of an excited state, it is known as *population inversion.** As mentioned earlier, the most important problem in a laser is to have stimulated emission dominating over spontaneous emission. For this to occur one must have population inversion and two methods for obtaining population inversion are discussed in Sections 22.1 and 22.5.

Let us next consider that the material in which population inversion has taken place is placed between two plane mirrors (see Fig. 22.7). As the beam bounces back and forth between the mirrors it gets amplified. However, since there will be some losses in each reflection one can obtain a threshold condition for the lasing action. If L_A represents the length of the active medium then in the absence of any losses the gain in traversing this distance would be $e^{\alpha_\nu L_A}$.

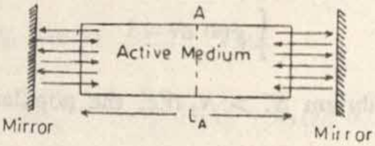


Fig. 22.7

Let I represents the intensity of the beam on one of the mirrors. In propagating to the other mirror, the intensity of the beam would become $Ie^{\alpha_\nu L_A}$. On reflection from the mirror of reflectivity R_1, the beam intensity would be $IR_1e^{\alpha_\nu L_A}$. Similarly in traversing the resonator length again and a further reflection at the mirror with reflectivity R_2, the beam intensity becomes $IR_1R_2e^{2\alpha_\nu L_A}$. In the steady state, one must have

$$IR_1R_2\,e^{2\alpha_\nu L_A}=I$$

or

$$R_1R_2\,e^{2\alpha_\nu L_A}=1$$

---

where m represents the mass of the molecule, T representing the absolute temperature. It may be mentioned that even in the absence of collisions or thermal motions, atomic lines are characterised by an intrinsic line width which represents the theoretical limit. This intrinsic line width which represents the theoretical frequency spread is $\sim \frac{1}{\tau}$ where τ represents the lifetime of the excited state.

*This is sometimes referred to as the 'negative temperature condition' because it is only at negative values of T, such 'population inversion' can occur at thermal equilibrium. However, it is not advisable to use this terminology of 'negative temperature' because neither the system corresponds to thermal equilibrium nor is there anything physical about negative temperature.

which is known as the threshold condition. For $R_1 R_2$ close to unity, $\alpha_\nu L_A$ would be small and we may write

$$1 + 2\alpha_\nu L_A = \frac{1}{R_1 R_2}$$

or

$$2\alpha_\nu L_A \approx 1 - R_1 R_2$$

But

$$\alpha_\nu = \kappa (N_1 - N_2) g(\nu)$$

Thus

$$N_1 - N_2 \approx \frac{[1 - R_1 R_2]}{\left[\frac{c^2}{\nu_0^2 \eta^2 8\pi} A\right]\left[\frac{2}{\Delta\nu}\sqrt{\frac{ln2}{\pi}}\right] 2L_A}$$

$$\approx \frac{2\pi^2 \Delta\nu (1 - R_1 R_2)}{\lambda^2 L_A \sqrt{\pi \, ln2} \, A}$$

where $\lambda = c/\nu_0\eta$; we have assumed $g(\nu)$ to correspond to the thermal distribution* and have taken the value of $g(\nu)$ at $\nu = \nu_0$, i.e. at the line centre. For $\lambda = 5 \times 10^{-5}$ cm, $\Delta\nu \approx 10^9$ sec^{-1}, $A \approx 10^7$ sec^{-1}, $R_1 R_2 \approx 0.99$, $L_A \approx 100$ cm, one obtains

$$N_1 - N_2 \approx 0.5 \times 10^8 \ \text{cm}^{-3}$$

22.4 Resonators

In the earlier sections we had seen how a light beam passing through a suitable medium with population inversion, may be amplified. In order to construct an oscillator which could supply light energy one must couple a part of the output into the medium. This can be achieved by placing the active medium between two mirrors which reflect most of the output energy back to the system. To obtain an output beam, one of the mirrors may be made partially reflecting. Such a system of two mirrors represents a resonant cavity. Thus, imagine a wave which starts from one of the mirrors and is travelling towards the other mirror. In passing through the active medium, it gets amplified. If the second mirror is partially reflecting then the wave is partially transmitted and the rest is reflected back towards the first mirror. In travelling to the first mirror, it again gets amplified and returns to the position it started from. Thus in between the two mirrors we have waves propagating along both directions. These waves interfere and in order that stable standing wave pattern be formed, the frequency must satisfy some conditions.

---

*The thermal effect gives rise to the Gaussian frequency distribution

$$g(\nu) = \frac{2}{\Delta\nu}\sqrt{\frac{ln2}{\pi}} \exp\left[-\frac{4\,ln\,2\,(\nu - \nu_0)^2}{(\Delta\nu)^2}\right]$$

Thus

$$g(\nu_0) = \frac{2}{\Delta\nu}\sqrt{\frac{ln2}{\pi}}$$

$$\left[\int g(\nu)\, d\nu = 1\right]$$

For the sake of simplicity, if we consider plane waves travelling to and fro, then the total phase change that the wave suffers after completing one complete round trip must be an integral multiple of 2π. Thus

$$\frac{2\pi}{\lambda}2L = 2m\pi; \quad m = 1, 2, 3, \ldots \tag{22.4-1}$$

where λ is the wavelength of the radiation in the medium enclosed by the cavity. If we put $\lambda = c/\nu$, then Eq. (22.4-1) becomes

$$\nu = m\frac{c}{2L} \tag{22.4-2}$$

which gives us the discrete frequencies in which the system will resonate. [Eq. (22.4-2) may be compared with Eq. (10.3-5) where we had obtained a similar result]. These are called the modes of oscillation of the resonant cavity and m represents the mode number. In obtaining the above relation, we have assumed that a plane wave can propagate unmodified inside the resonator. This is not true in general and the frequencies specified by Eq. (22.4-2) are only approximate. In general, the resonator oscillates not at a single frequency but at a continuum of frequencies distributed about each of the frequencies specified by Eq. (22.4-2). The widths of these lines are determined by the losses in the cavity which include absorption losses, losses due to partial reflection at the mirrors, diffraction losses of the beam, etc. When these losses are compensated by placing an active medium between the resonator mirrors, then the resultant laser emission is extremely narrow and is limited due to the presence of spontaneous emission and fluctuations in the cavity dimensions. The spontaneous emission results in line widths of only a fraction of a Hertz. In actual practice the fluctuations of the laser cavity are so large as to yield band widths of a few tens of kilohertz which is still an extremely monochromatic beam.

In Fig. 22.7 we have shown a simple resonator which is made up of two plane mirrors enclosing the active medium. The sides of the cavity are open so that radiation travelling at directions not nearly parallel to the axis of the resonator would escape from the resonator. This leads to an output beam which is also extremely directional. In general, the mirrors forming the cavity may be spherical. This helps in focussing the radiation so that one expects much lower diffraction losses which are caused due to the finite size of the cavity mirrors.

We will now consider a general resonator system bounded by two mirrors of radii of curvatures R_1 and R_2, separated by a distance d (see Fig. 22.8). We will assume that the medium between the two mirrors is homogeneous. We will use a ray optics analysis to obtain the condition for the resonator to be stable. In such resonators, a ray may keep bouncing back and forth and remain trapped in the resonator. This condition is termed as the stability condition. Further, we will be restricting ourselves to the paraxial region and will employ the matrix formulation which has been discussed in detail in Chapter 3.

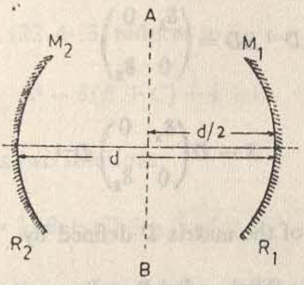


Fig. 22.8 A resonator consisting of two spherical mirrors.

Consider a ray which starts from a plane AB (which, for example, may be the plane midway between the mirrors, see Fig. 22.8) with coordinates x_0 and λ_0, where x_0 is the height of the ray and λ_0 is the optical direction cosine (see Sec. 3.2). This ray traverses a distance $d/2$ before intersecting the mirror M_1 and after suffering a reflection at M_1 it traverses a distance d before suffering another reflection at M_2. This reflected ray traverses another distance $d/2$ before returning to the plane AB. This constitutes one complete cycle of propagation of the ray. Let the matrix

$$S=\begin{pmatrix} B & -A \\ -D & C \end{pmatrix} \tag{22.4-3}$$

represent the matrix for one complete cycle of oscillation of the ray. Thus, if (x_1, λ_1) represents the coordinates of the ray after one cycle then we may write

$$\begin{pmatrix} \lambda_1 \\ x_1 \end{pmatrix}=\begin{pmatrix} B & -A \\ -D & C \end{pmatrix}\begin{pmatrix} \lambda_0 \\ x_0 \end{pmatrix} \tag{22.4-4}$$

For the resonator to be stable, we require that the ray must have the same coordinates (as it had when it had started) after traversing (the resonator) a finite number of times (say m). Thus, after the second trip, the ray would be specified by

$$\begin{pmatrix} \lambda_2 \\ x_2 \end{pmatrix}=\begin{pmatrix} B & -A \\ -D & C \end{pmatrix}\begin{pmatrix} \lambda_1 \\ x_1 \end{pmatrix}=\begin{pmatrix} B & -A \\ -D & C \end{pmatrix}^2\begin{pmatrix} \lambda_0 \\ x_0 \end{pmatrix} \tag{22.4-5}$$

Similarly after m traversals, the ray matrix would be

$$\begin{pmatrix} \lambda_m \\ x_m \end{pmatrix}=\begin{pmatrix} B & -A \\ -D & C \end{pmatrix}^m\begin{pmatrix} \lambda_0 \\ x_0 \end{pmatrix} \tag{22.4-6}$$

In order to determine the m th power of the matrix, we now diagonalise the matrix.

If S represents the system matrix, the matrix D is said to diagonalise S if

$$D^{-1}SD = \begin{pmatrix} \delta_1 & 0 \\ 0 & \delta_2 \end{pmatrix}$$

or

$$S = D\begin{pmatrix} \delta_1 & 0 \\ 0 & \delta_2 \end{pmatrix}D^{-1} \tag{22 4-7}$$

where D^{-1} is the inverse of the matrix D defined by

$$DD^{-1} = D^{-1}D = I \tag{22.4-8}$$

I, being the (2×2) unit matrix. Thus,

$$\begin{aligned} S^m &= \left[D\begin{pmatrix} \delta_1 & 0 \\ 0 & \delta_2 \end{pmatrix}D^{-1}\right]\left[D\begin{pmatrix} \delta_1 & 0 \\ 0 & \delta_2 \end{pmatrix}D^{-1}\right]\ldots m \text{ times} \\ &= D\begin{pmatrix} \delta_1^m & 0 \\ 0 & \delta_2^m \end{pmatrix}D^{-1} \end{aligned} \tag{22.4-9}$$

The quantities δ_1 and δ_2 are termed as the eigenvalues of the matrix. In order to determine the two eigenvalues we write

$$\begin{pmatrix} B & -A \\ -D & C \end{pmatrix}\begin{pmatrix} \xi \\ \eta \end{pmatrix} = \delta\begin{pmatrix} \xi \\ \eta \end{pmatrix} \tag{22.4-10}$$

Thus the effect of multiplying the column matrix (with components ξ and η) with the system matrix yields us the same column matrix multiplied by δ. The above equation can be rewritten as

$$\begin{pmatrix} B-\delta & -A \\ -D & C-\delta \end{pmatrix}\begin{pmatrix} \xi \\ \eta \end{pmatrix} = 0 \tag{22.4-11}$$

For the above equations to have non-trivial solutions, we must have

$$\det\begin{bmatrix} B-\delta & -A \\ -D & C-\delta \end{bmatrix} = 0 \tag{22.4-12}$$

i.e.

$$(B-\delta)(C-\delta) - AD = 0$$

or

$$\delta^2 - \delta(B+C) + (BC-AD) = 0 \tag{22.4-13}$$

If we use the fact that

$$BC - AD = 1 \tag{22.4-14}$$

(see Sec. 3.2), then Eq. (22.4-13) reduces to

$$\delta^2 - \delta(B + C) + 1 = 0 \tag{22.4-15}$$

The above equation has two solutions:

$$\delta = \tfrac{1}{2}(B + C) \pm \sqrt{\tfrac{1}{4}(B + C)^2 - 1} \tag{22.4-16}$$

From Eq. (22.4-6) it follows that for the ray to remain trapped in the cavity

$$\begin{pmatrix} B & -A \\ -D & C \end{pmatrix}^m \tag{22.4-17}$$

should not diverge as m increases. Thus, we must have

$$\delta_1 = e^{i\theta}, \ \delta_2 = e^{-i\theta} \tag{22.4-18}$$

where $i = \sqrt{-1}$ and θ is real. For δ to be a complex quantity, we see from Eq. (22.4-16) that the following inequalities must be satisfied

$$-2 \leqslant (B + C) \leqslant 2 \tag{22.4-19}$$

This is the *stability condition*. Given a resonator system, the system matrix representing a single traversal can be obtained, which would in turn give us the values of B and C. For the resonator system to be stable. Eq. (22.4-19) must be satisfied.

In order to determine an explicit expression for the stability condition, we consider a resonator system consisting of two mirrors of radii of curvatures R_1 and R_2 separated by a distance d (see Fig. 22.8). In a manner similar to that employed in Sec. 3.2, we can show that the effect of a mirror of radius of curvature R can be represented by the matrix

$$\begin{pmatrix} 1 & -2/R \\ 0 & 1 \end{pmatrix} \tag{22.4-20}$$

where the slope of the ray is always measured from the forward direction of travel. If the ray starts from the plane AB, then the system matrix for one complete round trip would be

$$\begin{pmatrix} B & -A \\ -D & C \end{pmatrix} = \begin{pmatrix} 1-\frac{d}{R_1}-\frac{3d}{R_2}+\frac{2d^2}{R_1R_2} & -\frac{2}{R_1}-\frac{2}{R_2}+\frac{4d}{R_1R_2} \\ -\frac{d}{2}-\frac{d^2}{2R_1}+\left(1-\frac{d}{R_2}\right)\left(\frac{3d}{2}-\frac{d^2}{R_1}\right) & -\frac{d}{R_1}+\left(1-\frac{d}{R_2}\right)\left(1-\frac{2d}{R_1}\right) \end{pmatrix} \tag{22.4-21}$$

Thus

$$B = 1-\frac{d}{R_1}-\frac{3d}{R_2}+\frac{2d^2}{R_1R_2} \tag{22.4-22}$$

and

$$C = 1-\frac{3d}{R_1}-\frac{d}{R_2}+\frac{2d^2}{R_1R_2} \tag{22.4-23}$$

Hence the stability condition [Eq. (22.4-19)] becomes

$$-2 \leqslant 2-\frac{4d}{R_1}-\frac{4d}{R_2}+\frac{4d^2}{R_1R_2} \leqslant 2 \tag{22.4-24}$$

which can also be written as

$$0 \leqslant \left(1-\frac{d}{R_1}\right)\left(1-\frac{d}{R_2}\right) \leqslant 1 \tag{22.4-25}$$

Hence, for the resonator to be stable, d, R_1 and R_2 must satisfy the above condition. Even though we have obtained the above result using ray arguments it can be shown to be applicable even when diffraction effects are also taken into account.

Fig. 22.9 depicts the stability diagram. The two coordinate axes correspond to $\left(1-\frac{d}{R_1}\right)$ and $\left(1-\frac{d}{R_2}\right)$. Only those resonator configurations that lie in the shaded region correspond to a stable configuration. The origin corresponds to $d = R_1 = R_2$. A resonator configuration with $d = R_1 = R_2$ is called confocal because the foci of both the mirrors coincide. The point marked A corresponds to $R_1 = R_2 = \infty$, i.e. the resonator is bounded by plane parallel mirrors. Such a resonator is termed a planar resonator. The point marked B corresponds to $d = 2R_1 = 2R_2$. In such a configuration the centres of curvature of both mirrors coincide and hence is called a concentric resonator system.

The above analysis has been made neglecting diffraction effects. In general, a wave which starts from one of the mirrors undergoes diffraction and looses some power due to spillover at the other mirror. The reflected wave undergoes further diffraction in travelling back to the first mirror. Thus, it is not possible for a homogeneous plane wave to resonate in the open resonator. A wave analysis of the resonator system yields the result that a wave with a Gaussian dependence of intensity will approximately resonate in the resonator. Thus the beem from most lasers is Gaussian. Since the intensity decreases rapidly away from the axis, the diffraction losses are also little.

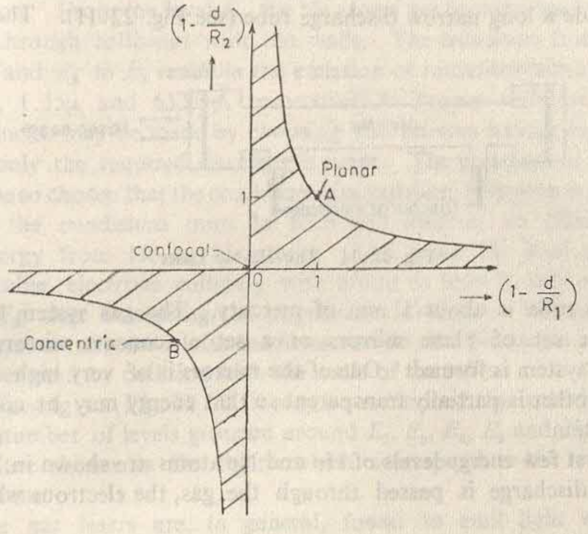

$\left(1-\frac{d}{R_2}\right)$
Planar
A
1
confocal
0
1
$\left(1-\frac{d}{R_1}\right)$
Concentric
B

Fig. 22.9 The stability diagram.

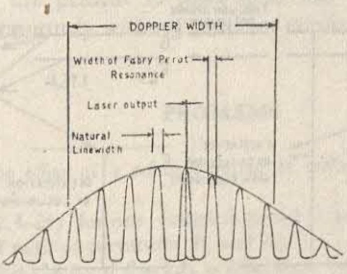


Fig. 22.10 The solid curve represents a typical Doppler broadened spectral line. The closely spaced cavity modes are shown as narrow peaks inside the curve. The sharp line represents the output of the laser [Adapted from Herriot (1962)].

In Fig. 22.10 the solid curve represents the Doppler width of the spectral line. The closely spaced cavity modes are shown as sharp peaks inside the curve. The width of these modes is primarily due to the losses in the cavity. When these losses are compensated, one obtains an extremely narrow line as shown in the figure.

22.5 Helium-Neon Laser

We will now briefly discuss the He-Ne laser which was first fabricated by Ali Javan and his coworkers at Bell Telephone Laboratories in USA. This was also the first gas laser to be operated successfully.

The He-Ne laser consists of a mixture of He and Ne of about 10: 1 placed inside a long narrow discharge tube (see Fig. 22.11). The pressure inside the tube is about 1 mm of mercury. The gas system is enclosed between a set of plane mirrors or a set of concave mirrors so that a resonator system is formed. One of the mirrors is of very high reflectivity while the other is partially transparent so that energy may be coupled out of the system.

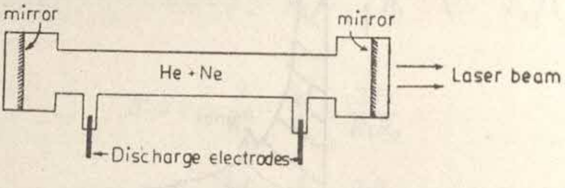


Fig. 22.11 The He-Ne Laser.

The first few energy levels of He and Ne atoms are shown in Fig 22.12. When a discharge is passed through the gas, the electrons which travel down the tube collide with the He atoms and excite them to the levels marked F_2 and F_3. These levels are metastable, i.e. He atoms excited to these states stay in these levels for a sufficiently long time before losing energy through collisions. Through these collisions, the Ne atoms are excited to the levels marked as E_4 and E_6 which have nearly the same energy as the levels F_2 and F_3 of He. Thus when He atoms in levels F_2 and F_3 collide with unexcited Ne atoms, they raise them to the levels E_4 and E_6 respectively. This results in a sizeable population of the levels

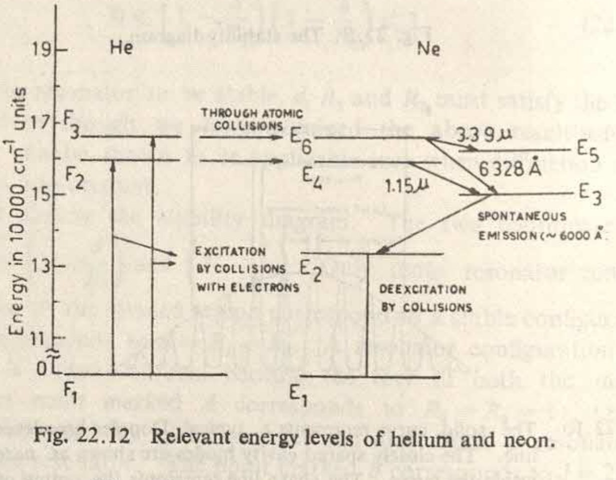

19
17
15
13
11
0
Energy in 10000 cm^{-1} units
He
Ne
F_3
F_2
F_1
THROUGH ATOMIC COLLISIONS
E_6
E_4
E_5
E_3
E_2
E_1
3.39 µ
6328 Å
1.15 µ
SPONTANEOUS EMISSION (~ 6000 Å
EXCITATION BY COLLISIONS WITH ELECTRONS
DEEXCITATION BY COLLISIONS

Fig. 22.12 Relevant energy levels of helium and neon.

E_4 and E_6. The population in these levels happens to be much more than those in the lower levels E_3 and E_5. Thus a state of population inversion has been achieved and any sponteneously emitted photon can trigger laser action in any of the three transitions shown in Fig. 22.12. The Ne atoms then drop down from the lower laser levels to the level E_2 through spontaneous emission. From the level E_2, the Ne atoms are brought back to the ground state through collisions with the walls. The transition from E_6 to E_5, E_4 to E_3 and E_6 to E_3 result in the emission of radiation having wavelengths 3.39μ, 1.15μ and 6328 Å respectively.* Proper selection of different frequencies may be made by choosing end mirrors having high reflectivity over only the required wavelength range. The pressures of the two gases must be so chosen that the condition of population inversion is not quenched. Thus the conditions must be such that there is an efficient transfer of energy from He to Ne atoms. Also since the level marked E_2 is metastable, electrons colliding with atoms in level E_2 may excite them to level E_3, thus decreasing the population inversion. The tube containing the gaseous mixture is also made narrow so that Ne atoms in level E_2 can get de-excited by collision with the walls of the tube.

Referring to Fig. 22.11, it may be mentioned that actually there are a large number of levels grouped around E_2, E_3, E_4, E_5 and E_6. Only those levels are shown in the figure which correspond to the important laser transitions.

The gas lasers are, in general, found to emit light which is more directional and more monochromatic. This is because of the absence of the effects such as crystalline imperfection, thermal distortion and scattering which are present in solid-state lasers. Gas lasers are capable of operating continuously without need for cooling.

PROBLEMS

1. Show that the effect of a concave mirror of radius of curvature R is given by Eq. (22.4-20).
2. Derive Eq. (22.4-21) for one complete round trip of a resonator bounded by two mirrors of radii of curvatures R_1 and R_2.
3. Referring to Fig. 22.8, if $d = 2R_1 = 2R_2$, show that all rays passing through the common centre of curvature of the mirrors will retrace their path and hence be trapped inside the cavity. What will be the ray paths for $d = R_1 = R_2$?

REFERENCES AND SUGGESTED READING

ALLEN, L. (1969), *Essentials of Lasers*, Pergamon Press, Oxford.

---

*Notice that the laser transitions corresponding to 3.39μ and 1.15μ are not in the visible region. The 6328 Å transition corresponds to the well known red light of the He-Ne laser.

BROWN, R. (1969), *Lasers: A Survey of Their Performance and Applications*, Business Books, London.

GHATAK, A.K. and LOKANATHAN, S. (1975), *Quantum Mechanics*, Macmillan, New Delhi.

HEAVENS, O.S. (1971), *Lasers*, Duckworth, London.

HERRIOT, D.R. (1962), Optical properties of a continuous He-Ne Optical Maser, *J. Opt. Soc. Am.* (USA), **52**. 31.

LENGYL, B.A. (1966), *Introduction to Laser Physics*, John Wiley, New York.

SCHAWLOW, A. (1969), *Lasers and Light* (A collection of Scientific American articles) W.H. Freeman and Co., San Francisco).

SIEGMAN, A.E. (1971), *An Introduction to Lasers and Masers*, McGraw-Hill Book Co., New York.

SOBOLEV, N. (1974), *Lasers and their Prospects*, Mir Publishers, Moscow.

SODHA, M.S. and GHATAK A.K. (1977), *Inhomogeneous Optical Waveguides*, Plenum, New York.

. And can I ever bid these joys farewell ?

KEATS

General Books on Optics

ANDREWS, C.L. (1960), *Optics of the Electromagnetic Spectrum*, Prentice-Hall, Englewood Cliffs, New Jersey (USA).

BARTON, A.W. (1939), *A Textbook on Light*, Longmans, Green & Co., London (England).

BLAKER, J.W. (1971), *Geometric Optics, The Matrix Theory*, Marcel Dekker Inc., New York (USA).

BORN, M. AND WOLF, E. (1975), *Principles of Optics*, Pergamon Press, Oxford (England).

BROWN, E.B. (1965), *Modern Optics*, Reinhold, New York (USA).

CURRY, C. (1957), *Wave Optics*, Edward Arnold, London (England).

DITCHBURN, R.W. (1963), *Light*, John Wiley & Sons, New York (USA).

DRUDE, P. (1939), *The Theory of Optics*, Longmans, Green & Co., London (England).

FRANCON, M., KRAUZMAN, N., MATHIEU, J.P. AND MAY, P. (1970), *Experiments in Physical Optics*, Gordon and Breach, New York (USA).

FRANK, N.H. (1950), *Introduction to Electricity and Optics*, McGraw-Hill Book Co., New York (USA).

FRY, G.A. (1969), *Geometrical Optics*, Chilton, Philadelphia (USA).

GARBUNY, M. (1965), *Optical Physics*, Academic Press, New York (USA).

GOLDWASSER, E.L. (1965), *Optics, Waves, Atoms and Nuclei: An Introduction*, Benjamin, New York (USA).

HARDY, A.C. AND PERRIN, F.H. (1932), *The Principles of Optics*, McGraw-Hill Book Co., New York (USA).

HECHT, E. AND ZAJAC, A. (1974), *Optics*, Addison Wesley Publishing Co., Reading, Mass. (USA).

HOUSTON, R.A. (1938), *A Treatise on Light*, Longmans, Green & Co., London (England).

HUYGENS, C. (1962), *Treatise on Light*, Dover Publications, New York (USA).

JENKINS, F.A. AND WHITE, H.E. (1976), *Fundamentals of Optics*, McGraw-Hill Book Co., New York (USA).

Johnson, B.K. (1947), *Optics and Optical Instruments*, Dover Publications, New York (USA).

Klein, M.V. (1970), *Optics*, John Wiley & Sons, New York (USA).

Lipson, S.G. and Lipson, H. (1969), *Optical Physics*, Cambridge University Press, London (England).

Longhurst, R.S. (1967), *Geometrical and Physical Optics*, John Wiley & Sons, New York (USA).

Mach, E. (1926), *The Principles of Physical Optics, An Historical and Philosophical Treatment*, Dover Publications, New York (USA).

Martin, L.C. (1955), *Geometrical Optics*, Pitman, London (England).

Mathieu, J.P. (1975), *Optics*, Pergamon Press, Oxford (England).

Mathur, B.K. (1969), *Principles of Optics*, Gopal Press, Kanpur.

Mayer-Ardent, J.R. (1972), *Introduction to Classical and Modern Optics*, Prentice-Hall, Englewood Cliffs, New Jersey (USA).

Morgan, J. (1953), *Introduction to Geometrical and Physical Optics*, McGraw-Hill Book Co., New York (USA).

Newton, I. (1952), *Optiks*, Dover Publications, New York (USA).

Noakes, G.R. (1944), *A Text Book of Light*, Macmillan, London (England).

Nussbaum, A. (1968), *Geometrical Optics: An Introduction*, Addison Wesley, Reading, Mass. (USA).

Palmer, C.H. (1962), *Optics, Experiments and Demonstrations*, Johns Hopkins Press, Baltimore (USA).

Robertson, J.K. (1957), *Introduction to Optics, Geometrical and Physical*, Van Nostrand, Princeton, New Jersey (USA).

Ronchi, J. (1971), *The Nature of Light*, Harvard University Press, Cambridge, Mass. (USA).

Rossi, B. (1957), *Optics*, Addison Wesley Publishing Co., Reading, Mass. (USA).

Sears, F.W. (1944), *Optics*, Addison Wesley Publishing Co., Reading, Mass. (USA).

Simmons, J.W. and Guttmann, M.J. (1970), *States, Waves and Photons: A Modern Introduction to Light*, Addison Wesley Publishing Co., Reading, Mass. (USA).

Sommerfeld, A. (1964), *Optics*, Academic Press, New York (USA).

Stone, J.M. (1963), *Radiation and Optics*, McGraw-Hill Book Co., New York (USA).

Strong, J. (1958), *Concepts of Classical Optics*, Freeman, San Francisco (USA).

Tenquist, D.W., Whittle, R.M. and Yarwood, J. (1970), *University Optics* (2 volumes) Iliffe, London (England).

Valasek, J. (1949), *Optics: Theoretical and Experimental*, John Wiley & Sons, New York (USA).

Van Heel, A.C.S. and Velzel, C.H.F. (1968), *What is Light*? Weidenfeld and Nicholson, London (England).

Wagner, A.F. (1929), *Experimental Optics*, John Wiley & Sons, New York (USA).

Webb, R.H. (1969), *Elementary Wave Optics*, Academic Press, New York (USA).

Wood, R.W. (1934), *Physical Optics*, Dover Publications, New York (USA).

Young, H.D. (1968), *Fundamentals of Optics and Modern Physics*, McGraw-Hill Book Co., New York (USA).

Author Index

Aggarwal, A.K., 29, 44
Allen, L., 526
Alonso, M., 409, 439
Andrews, C.L., 528
Arons, A.B., 176, 187

Baker, A., 17, 229, 279
Ball, C.J., 383
Ballard, S.S., 407, 408, 439
Balmain, K.G., 489
Barton, A.W., 4, 17, 528
Baumeister, P., 279
Beams, J.W., 497, 503
Bird, G.R., 407, 439
Blaker, J., 64, 528
Born, M., 10, 17, 22, 44, 88, 279, 292, 383, 439, 508, 528
Bose, G., Plate 8 (facing p. 255)
Bottcher, C.J.F., 117, 118, 143
Braddick, H.J., 143, 169, 187, 383
Bragg, L., 383
Brouwer, W., 64
Brown, E.B., 528
Brown, J.C., 397
Brown, R., 508, 526
Bryngdahl, O., 480, 489
Burch, J.M., 64
Burckhardt, C.B., 398
Bush, R.T., 44

Cagnet, M., 89, 279
Caulfield, H.J., 397
Cave, E.F., 279
Chandrasekhar, S., 439
Chiao, R.V., 341, 383
Cleveland, E.L., 279
Collett, E., 279
Collier, R.J., 398
Compton, A.H., 500, 501, 502
Cook, A.H., 279
Corson, D.R., 463, 489
Coulson, C.A., 169
Cowley, J.M., 383
Crawford, F.S., 143, 169
Curry, C., 528

Davis, S.P., 64
de Broglie, L., 17, 503
DeWitte, H.A., 187
Dirac, P.A.M., 12, 17
Ditchburn, R.W., 194, 279, 528
Drude, P., 528

Eakin, D.M., 64
Elmore, W.C., 169, 383

Feinberg, G., 17
Feynman, R.P., 1, 12, 14, 17, 44, 143, 229, 279, 416, 439, 483, 489.

Finn, E.J., 409, 439
Flint, H.T., 396, 398
Forrester, A.T., 296, 298
Francon, M., 89, 279, 298, 383, 528
Frank, N.H., 169, 528
Fraser, A., 29, 44
French, A.P., 143
Fry, G.A., 528
Friedmann, G.B., 489

Gamow, G., 8, 17
Garbuny, M., 528
Garmire, E., 341, 383
Gay, P., 439
Gerrard, A., 64
Ghatak, A.K., 29, 30, 35, 43, 44, 338, 339, 341, 384, 466, 490, 508, 512, 526
Givens, M.P., 398
Goldwasser, E.L., 528
Goodman, J.W., 344, 383
Gudmundsen, R.A., 296, 298
Gupta K.K., Plate 2 (facing p. 72) Plate 17 (facing p. 379)
Guttmann, M.J., 18, 64, 529

Halbach, K., 64
Hardy, A.C., 528
Harte, J.A., 397
Haus, H.A., 339, 384
Heald, M.A., 169, 383
Heavens, O.S., 526
Hecht, E., 528
Heirtzler, J.R., 489
Heisenberg, W., 17
Herriot, D.R., 523, 526
Hoffman, B., 503
Holroyd, L.V., 279
Hopkins, H.H., 89
Houston, R.A., 528
Humphreys, W.J., 44
Huygens, C., 187, 528

Jaffe, J.H., 279
Jaseja, T.S., 286, 298
Javan, A., 286, 298
Jenkins, F.A., 176, 279, 383, 424, 439, 529, Plate 9 (facing p. 272)
Johnson, B.K., 529
Johnson, P.O., 296, 298
Joos, G., 157, 169
Jordan, E.C., 489
Joyce, A., 2, 3, 18
Joyce, W.B., 2, 3, 18

Kaw, P.K., 29, 44
Khular, E., 29, 30, 43, 44
Klein, M.V., 529
Krauzman, N., 528

Lawrence, E.O., 497, 503
Leighton, R.B., 1, 17, 44, 143, 229, 279 439, 489
Leith, A.N., 398
Lengyl, B.A., 508, 526
Lin, L.H., 398
Lipsett, M.S., 296, 297, 298
Lipson, H., 529
Lipson, S.G., 529
Lokanathan, S., 512, 526
Longhurst, R.S., 529
Lorrain, P., 463, 489
Loudon, R., 117, 143
Lu, S., 397

Mach, E., 529
Mach, W.H., 29, 44
Mallick, S., 279
Mandel, L., 13, 18, 296, 297, 298
Marburger, J.M., 339, 384
Martin, L.C., 529
Mathieu, J.P., 528, 529
Mathur, B.K., 529
May, P., 528
Mayer-Ardent, J.R., 529
Meiners, H.F., 298, 383, Plate 3 (facing p. 83), Plate 10 (facing p. 309), Plate 13 (facing p. 319), Plate 14 (facing p. 324)
Methernal, A.F., 398
Milford, F.J., 489
Morgan, J., 529

Newton, I., 529
Noakes, G.R., 529
Nussbaum, A., 64, 529

Oppenheim, V., 279
Oppenheimer, J.R., 18

Pain, H.J., 143, 169
Palmer, C.H., 529
Panofsky, W.K.H., 339, 383, 451, 489
Parrish, M.P., 407, 439
Penfield, R.H., 44
Pennington, K.S., 398
Perrin, F.H., 528
Phillips, M., 339, 383, 451, 489
Phillips, R.A., 384, Plate 15 (facing p. 346)
Pincus, G., 279

Reitz, J.R., 489
Resnick, R., 143
Robertson, J.K., 529
Robinson, R.S., 44
Ronchi, J., 529
Rose, A., 18, Plate 1 (facing p. 11)
Rossi, B., 529

Sandhu, H.S., 489
Sands, M., 1, 17, 44, 143, 229, 279, 439, 489.
Schawlow, A., 508, 526
Schilpp, P.A., 18
Schull, C.G., 10, 18
Schurcliff, W.A., 407. 408, 439
Sears, F.W., 529
Shankland, R.S., 500, 503
Sharma, A., Plate 11 (facing p. 311), Plate 12 (facing p. 315)
Shaw. J.H., 44
Siegman, A.E., 508, 526
Simmons, J.W., 18, 64, 529
Sirohi, R.S., Plate 6 (facing p. 221), Plate 7 (facing p. 253), Plate 18 (facing p. 386), Plate 19 (facing p. 397)
Sladkova, J., 279
Slater, J.C, 169
Smith, C.J., 74, 89
Smith, H.M., 398
Sobolev, N., 508, 526
Sodha, M.S., 29, 35, 44, 338, 339, 341, 384, 466, 490, 508, 526
Sommerfeld, A., 113, 119, 143, 529
Steel, W.H., 279
Stone, J.M., 143, 384, 529
Strong. J., 529

Taylor, G.I., 11
Tenquist. D.W., 529
Thierr, J.C., 89, 279
Thyagarajan, K., 29, 30, 35, 43, 44, Plate 16 (facing p. 366)
Tolansky, S., 279
Townes, C.H., 286, 298, 341, 383
Tripathi, V.K., 338, 339, 341, 384

Upatnieks, J., 398

Valasek, J., 529
Van Heel, A.C.S., 529
Velzel, C.H.F., 529
Venkateshwarulu, D., 398

Wagner, A.F., 529
Wagner, W.G., 339, 384
Waldron, R.A., 169
Waterman, T.H., 439
Webb, R.H., 529
Welford, W.T., 89
Weisskopf, V.W., 503
Weltinh, H., 298
White, H.E, 176, 279, 383, 424, 439 529,
Plate 9 (facing p. 272)
Whittle, R.M., 529
Wichmann, E.H., 18, 503
Wolf, E., 13, 18, 22, 44, 88, 279, 292, 383, 439, 528
Wood, E.A., 439
Wood, R.W., 529
Worsnop, B.L., 396, 398

Yarwood, J., 529
Young, H.D., 529

Zajac, A., 528

Subject Index

Abbe-sine condition, 84
Aberrations, 65, 70
Achromatic doublet, 67
Active medium, 516
Airy pattern, 311, 381
Ampere's law, 463
Analyser, 404
Anisotropic medium
  path of rays in, 37
  plane wave propagation in, 430
Antinodes, 193
Aplanatic points, 74
Astigmatism, 84

Barrel distortion, 88
Beats, 204, 293
  optical, 293
Biaxial crystal, 435
Birefringence (*see* double refraction)
Blooming, 242
Body centred cubic lattice, 332
Bragg's law, 334
Brewster angle (*see* polarizing angle)

Calcite crystal, 419
Cauchy relation, 113
Chromatic aberration, 66
Chromium-ion, energy levels of, 509
Circular aperture,
  Fraunhofer diffraction by, 311, 379
  Fresnel diffraction by, 350, 373
Circularly polarized wave, 402, 415
Coefficient of Finesse, 269
Coherence, 212, 280
Coherence length, 260
Coherent sources, 213, 288
Collisional broadening, 515
Colours of thin films, 253
Coma, 76, 81
Compensating plate, 262
Complex representation, 203
Compton effect, 497
Compton scattering, 498
Compton shift, 502
Conducting medium,
  reflection by, 486
  wave equation in, 457
Constitutive relations, 441
Continuity conditions, 459
Continuity equation, 450, 463
Copper, x-ray diffraction pattern of, 338, 348
Cornu's spiral, 364
Corpuscular Model, 2
  (*see* also Photon)
Cosine law, 238
Crystal structure, 331

Curvature of field, 84

Damped simple harmonic motion, 103
de-Broglie relation, 10
Debye-Scherrer rings, 337
Delta function, 142, 344
Dichroism, 409
Dielectric constant, 112
  of ionic crystals, 119
  of metals, 118, 119
Dielectric permittivity, 430
Diffraction, Fraunhofer, 301
  Fresnel, 349
Diffraction grating, 325
  resolving power of, 329
Diffraction phenomenon, 301, 349
  explanation by Uncertainty Principle, 8
Diffused reflection, 178
Dipole (*see* oscillating dipole)
Dirac delta function, 142, 344
Dispersion electrons, 112
Dispersion relation, 138
Displacement current, 464
Displacement of fringes, 227
Distortion, 87
Double Refraction, 401, 409, 418

Einstein coefficients, 511
Einstein's Photon theory, 497
Electric susceptibility, 112
Electromagnetic spectrum, 447
Electromagnetic waves, 440
  energy density of, 449, 453
  intensity of, 453
  propagation in an anisotropic medium, 430
  propagation in a conducting medium, 457
  propagation in a dielectric, 441
  reflection of, 467
  refraction of, 467
  standing, 194
  transverse character of, 405, 442
Electron wave, 141
Electrostriction, 338
Ellipsoidal reflector, 27
Elliptically polarized wave, 416, 417
Energy density of electromagnetic waves, 449, 453
Equation of continuity, 450, 463
Evanescent waves, 480
Extraordinary rays, 37, 419

Fabry-Perot interferometer, 271
  resolving power of, 272
Face-centred cubic lattice, 333
Fermat's principle, 21, 23
Finesse, coefficient of, 269
Forced vibrations, 103, 130
Fourier analysis, 121
Fourier integral, 132
Fourier series, 121
Fourier transform plane, 343, 346
Fraunhofer diffraction, 301
  by a circular aperture, 311
  by a grating, 326
  by a single slit, 303
  by two slits, 317
Frequency, 147
Fresnel biprism, 224
Fresnel diffraction, 349
  by a circular aperture, 350
  by a long narrow slit, 368
  by an opaque disc, 354, 376
  by a rectangular aperture, 378
  by a straight edge, 358
Fresnel equations, 483
Fresnel half period zones, 350
Fresnel integrals, 363
  table of, 364
Fresnel's two mirror arrangement, 223
Fringes of equal thickness, 249
Fringe visibility in the double hole experiment, 286
Fringe width formula, 217, 225
Fuchtbauer-Landenburge formula, 515

Gaussian wavepacket, 138, 167
  propagation in a dispersive medium, 138, 140, 142, 143
Graphical method for studying superposition of waves, 200
Grating spectrum, 326
  for oblique incidence, 330
  (*see* also Diffraction grating)
Grazing incidence, 474
Group velocity, 134

Half period zones, 350
Half wave plate, 428
Helical ray, 43
Helium, energy levels of, 524
Helium-neon laser, 524
Hertz's experiment, 6
High reflectivity by thin film deposition, 242
Hologram,
  formation of, 386, 390
  reconstruction of, 386, 390
Holography, 385
  applications of, 394

physical significance of, 461
Huygens' eye piece, 80
Huygens' principle, 170
  application in inhomogeneous media, 185
  application in refraction and reflection, 174
Hydrogen atom, ionisation of energy of, 17

Incoherent sources, 213, 220, 288
Inhomogeneous media, 29, 185
  ray paths in, 29
Initial phase, 95
Intensity of electromagnetic waves, 453
Interference by a wedge, 247
Interference of polarized light, 424
Interference filters, 277
Interference of two beams,
  by division of amplitude, 234
  by division of wavefront, 207
Interference with water waves, 207
Interfersnce with white light, 226
Inverse square law, 150
Ives' experiment, 197

Kerr effect, 338

Lasers, 507
Lateral spherical aberration, 71
Laue spots, 336, 337
Laws of reflection, 23, 177, 467
Laws of refraction, 23, 174, 467
  at the surface of an anisotropic medium, 37, 43
Left circularly polarized wave, 416
Lens law, 57
Limit of resolution, 313
Linearly polarized wave, 402, 413
Line width, 282, 523
Lloyd's mirror arrangement, 228, 231
Longitudinal sound waves,
  through a gas, 158
  through a solid, 155
Longitudinal spherical aberration, 71
Looming, 31, 32
Lorentzian pulse, 147
Lummer-Gehrcke plate, 275

Malus' law, 412
Marginal rays, 71
Matrix method, 45, 519
Matter waves, 10
  probabilistic interpretation of, 10
Maxwell's equations, 440
Metastable state, 509
Michelson interferometer 12, 260, 282
Michelson's Steller Interferometer, 292
Mirage, 29
Modes of a resonator, 518
Multiple beam interferometry, 266
Multiple slit Fraunhofer diffraction pattern, 321
  actual intensity distribution, 324
  positions of maxima and minima, 322

Negative crystal, 418
Neon, energy levels of, 524
Newton's particle model, 2
Newton's rings, 253
Nicol prism, 410
Nodal curves, 210
Nodal planes, 58
Nodes, 193
Nonlinear self-focusing, 338
Non-reflecting films, 229

Oblique incidence, 330
One dimensional wave equation, 152
  solution of, 160
Optical beats, 293
Optical waveguide, 465, 466
Optic axis, 37, 418
Ordinary ray, 419
Oscillating dipole, 451

Paraboloidal reflector, 26
Paraxial optics, 45
Partially polarized wave, 429, 473
Particle nature of radiation, 7, 493
Phase, 95
Phase change on reflection, 229, 484
Phase difference, 95
Phase velocity, 137
Photoelectric effect, 493
Photon, 497
  polarization of, 14
Pin-cushion distortion, 88
Planck's law, 511
Plane waves in an anisotropic medium, 430
Plane waves in a dielectric, 441
Plasma frequency, 114, 120
Plucked string, transverse vibrations of, 126
Pohl's interferometer, 232, 243
Polarization, 401
Polarization by double refraction, 409
Polarization by reflection, 408
Polarization by scattering, 410

Polarized Light,
  requirements of, 394
  analysis of, 429
  interference of, 424
  production of, 406
Polarizing angle, 409, 473, 484
Polaroid, 407, 408
Polycrystalline material, x-ray diffraction of, 337
Population inversion, 512
Positive crystal, 418
Powder method, 337
Poynting vector, 432, 449, 471
Principal dielectric constants, 430
Principal maxima, 323
  width of, 324
Pulse dispersion, 138, 140
Pulse propagation, 145

Quarter wave plate, 427
Quasi-conductor, 458

Radiation,
  particle nature of, 493
Radiation pressure, 454
  measurement of, 456
Ray, 22
Ray equation, 34
Rayleigh criterion, 315
Ray refractive index, 432
Reconstruction process, 386
Rectangular aperture, Fresnel diffraction by, 378
Rectilinear propagation, 173
Reflection by a conducting medium, 486
Reflection coefficient, 471, 472, 476
Reflection laws, 23, 467
  using Huygens' principle, 177
Reflection of electromagnetic waves, 467
Refraction of electromagnetic waves, 467
Refraction laws, 23, 467
  using Huygens' principle, 174
Refraction matrix, 49
Refraction at a spherical surface, 50
Refraction of a spherical wave,
  by a plane surface, 179
  by a spherical surface, 179
Refractive index, 112
  origin of, 111
  relation to dielectric permittivity, 112, 445
Refractive indices, tables of, 176
Resolying power,
  of a grating, 329
  of a microscope objective, 316
  of a telescope, 313
Resonance, 108
Resonators, 517
Right circularly polarized wave, 416
Ruby laser, 509

Sagittal focus, 86
Self-focusing phenomenon, 338
SELFOC fibre (*see* square law medium)
Separated doublet, 69
Shape factor of a lens, 76
Shock wave fronts, 184
Simple cubic lattice, 332
Simple harmonic motion, 93
Single slit Fraunhofer diffraction pattern, 303
  actual intensity distribution, 306
  positions of maxima and minima, 305
  transition from the Fresnel region, 372
Sinusoidal waves, 147
Snell's law, 3, 469
Sodium, x-ray diffraction pattern of, 338
Spatial coherence, 289
Spatial frequency, 346
Spatial frequency filtering, 343
Spherical aberration, 70
  of a plane refracting surface, 72
  of a spherical refracting surface, 74
  of a thin lens, 75
Spherical surface,
  refraction at, 50
Spherical waves, 167
Spontaneous emission, 508, 510
Square law medium, 34, 43, 185
  ray paths in, 35
Stability condition, 521
Stability diagram, 523
Standardisation of meter, 262
Standing electromagnetic waves, 194
Stationary light waves, 197
Stationary waves on a string, 192
Stimulated emission, 508, 510
Stokes' relations, 230, 240, 268, 474
Straight edge diffraction pattern,
  elementary theory, 358
  rigorous theory, 360
Stretched string, vibrations of, 100
Superposition of sinusoidal waves, 198
Superposition of two disturbances, 412
Superposition of waves, 191
Surface waves, 480
Susceptibility, 112

Tangential focus, 86
Taylor's experiment, 11

Temporal coherence, 280
Thick lens, 54
Thin films, colours of, 253
Thin lens formula, 55
  using Huygens' principle, 182
Three dimensional wave equation, 164, 447
Time energy uncertainty principle, 17
Total internal reflection, 177, 475
Translation matrix, 48
Transmission coefficient, 471, 472, 476
Transverse character of electromagnetic waves, 405, 442
Transverse electric waves, 465
Transverse magnetic waves, 465
Transverse vibrations, 126 154
Transverse vibrations of a plucked string, 126
Transverse waves, 150
Turning point, 187
Two slit Fraunhofer diffraction pattern, 317
  actual intensity distributions, 320
  positions of maxima and minima, 319

Uncertainty principle, 8
Uniaxial crystals, 418, 435
Unit planes, 55
Unpolarized wave, 404

Visibility, 288

Water waves, 150
Wave equation, one dimensional, 152
  three dimensional, 164, 447
Waveguide, 465, 466
Wavelength, 147
Wave model, 3
Wave packet, 134
Wave particle duality, 10
Wave propagation, 144, 430, 441, 457
Weiner's experiment, 197
White light interference, 226
Wire-grid polarizer, 407

X-ray diffraction, 331
  of copper, 338, 348
  of sodium, 337, 338
  of sodium chloride, 348

Young's double hole experiment, 13, 290
Young's interference pattern, 216
  the intensity distribution, 218

Zone plate, 355

S. C. E. R. T., West Bengal
Library
Calcutta
25/3, B. C. Road

Optics

This book covers both traditional aspects like ray tracing, interference, diffraction, polarization, reflection, refraction and propagation of electromagnetic waves, as well as some of the modern aspects of optics like matrix method, coherence, holography, self-focussing, spatial frequency filtering, lasers, etc.

Beginning with a brief historical introduction to the subject and a discussion on the nature of light, the text is divided in seven parts:

- Part I deals with geometrical optics including chapters on Fermat's principle, the matrix method and a discussion on aberrations;
- Part II deals with various types of vibrations and propagation of waves;
- Parts III and IV deals with interference and diffraction of waves;
- Part V discusses the electromagnetic theory of light, the phenomena of polarization, doubles refraction, reflection and refraction of waves;
- Part VI deals with the corpuscular nature of radiation; and in
- Part VII the physics of the laser has been discussed.

A large number of problems as well as references for additional reading are appended to each chapter.

This book was developed from lectures delivered by the author at the Indian Institute of Technology, Delhi, and is suitable for undergraduate and postgraduate students of physics. Students of electrical engineering, mechanical engineering as well as those in research cells of industries involved in the fabrication and application of lasers would also find the book useful.

Ajoy Ghatak

is currently Professor in the Department of Physics at the Indian Institute of Technology, Delhi. He obtained his MSc from Delhi University and PhD from Cornell University. After obtaining his PhD he worked as a research associate at the Brookhaven National Laboratory.

Dr Ghatak has published over 100 research papers in various areas of Physics; his current interests are in the fields of electromagnetic wave propagation and applied optics. He has co-authored *Quantum Mechanics* (Macmillan, New Delhi), *Inhomogeneous Optical Waveguides* (Plenum Press, New York), *An Introduction to Lattice Dynamics* (Addison Wesley, Reading) and *Self Focusing of Laser Beams in Dielectrics Plasmas and Semiconductors* (Tata McGraw-Hill, New Delhi).

Tata McGraw-Hill Publishing Company Limited
4/12 Asaf Ali Road, New Delhi 110 002

0-07-096421-1